



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:21:08 1999; MasPar time 3.41 Seconds
Tabular output not generated. -206.077 Million cell updates/sec

Title: >US-09-209-961-7
Description: (1-59) from US09209961.pep
Sequence: 1 MKCTLVFQSRFCIFPLTFKS.....TRIPLSNKVLTAVDRLRCP 59

Scoring table: PAM 150
Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1

Statistics: Mean 24.207; Variance 97.779; scale 0.248

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	77	17.3	91	2	US-08-421- Sequence 8, Applicatio	1.77e+01
2	77	17.3	91	2	US-08-467- Sequence 5, Applicatio	1.77e+01
3	77	17.3	91	2	US-08-660- Sequence 21, Applicati	1.77e+01
4	77	17.3	91	1	US-08-480- Sequence 21, Applicati	1.77e+01
5	77	17.3	91	1	US-08-347- Sequence 12, Applicati	1.77e+01
6	77	17.3	91	2	US-08-798- Sequence 12, Applicati	1.77e+01
7	77	17.3	91	1	US-08-375- Sequence 5, Applicatio	1.77e+01
8	77	17.3	91	2	US-08-633- Sequence 3, Applicatio	1.77e+01
9	77	17.3	263	3	PCT-US92-0 Sequence 4, Applicatio	1.77e+01
10	77	17.3	263	2	US-08-646- Sequence 4, Applicatio	1.77e+01
11	77	17.3	263	1	US-07-988- Sequence 4, Applicatio	1.77e+01
12	77	17.3	263	1	US-08-425- Sequence 4, Applicatio	1.77e+01
13	77	17.3	263	1	US-07-901- Sequence 4, Applicatio	1.77e+01
14	77	17.3	263	2	US-08-477- Sequence 4, Applicatio	1.77e+01
15	77	17.3	263	1	US-08-488- Sequence 4, Applicatio	1.77e+01
16	77	17.3	286	1	US-08-324- Sequence 13, Applicati	1.77e+01
17	76	17.1	326	2	US-08-560- Sequence 19, Applicati	2.13e+01
18	76	17.1	326	1	US-08-349- Sequence 19, Applicati	2.13e+01
19	76	17.1	326	1	US-07-918- Sequence 6, Applicatio	2.13e+01
20	76	17.1	326	1	US-08-233- Sequence 19, Applicati	2.13e+01
21	76	17.1	648	1	US-08-321- Sequence 4, Applicatio	2.13e+01
22	76	17.1	679	2	US-08-452- Sequence 15, Applicati	2.13e+01
23	76	17.1	679	1	US-07-998- Sequence 15, Applicati	2.13e+01

24	76	17.1	679	3	PCT-US92-1	Sequence 15, Applicati	2.13e+01
25	74	16.6	942	2	US-08-884- Sequence 4, Applicatio	Sequence 4, Applicatio	3.09e+01
26	72	16.2	345	1	US-08-403- Sequence 2, Applicatio	Sequence 2, Applicatio	4.46e+01
27	71	16.0	113	1	US-08-264- Sequence 10, Applicati	Sequence 10, Applicati	5.35e+01
28	71	16.0	314	1	US-08-118- Sequence 8, Applicatio	Sequence 8, Applicatio	5.35e+01
29	71	16.0	331	2	US-08-466- Sequence 166, Applicat	Sequence 166, Applicat	5.35e+01
30	71	16.0	331	2	US-08-464- Sequence 166, Applicat	Sequence 166, Applicat	5.35e+01
31	71	16.0	337	2	US-08-485- Sequence 162, Applicat	Sequence 162, Applicat	5.35e+01
32	71	16.0	337	2	US-08-464- Sequence 162, Applicat	Sequence 162, Applicat	5.35e+01
33	71	16.0	337	2	US-08-461- Sequence 162, Applicat	Sequence 162, Applicat	5.35e+01
34	71	16.0	459	2	US-08-638- Sequence 39, Applicati	Sequence 39, Applicati	5.35e+01
35	71	16.0	2873	3	PCT-US95-0	Sequence 15, Applicati	5.35e+01
36	71	16.0	2873	2	US-08-638- Sequence 2, Applicatio	Sequence 2, Applicatio	5.35e+01
37	71	16.0	2873	2	US-08-461- Sequence 15, Applicati	Sequence 15, Applicati	5.35e+01
38	71	16.0	2873	2	US-08-444- Sequence 15, Applicati	Sequence 15, Applicati	5.35e+01
39	71	16.0	2873	2	US-08-464- Sequence 15, Applicati	Sequence 15, Applicati	5.35e+01
40	71	16.0	2873	2	US-08-485- Sequence 15, Applicati	Sequence 15, Applicati	5.35e+01
41	71	16.0	2910	3	PCT-US95-0	Sequence 157, Applicat	5.35e+01
42	71	16.0	2910	2	US-08-464- Sequence 183, Applicat	Sequence 183, Applicat	5.35e+01
43	71	16.0	2910	2	US-08-444- Sequence 183, Applicat	Sequence 183, Applicat	5.35e+01
44	71	16.0	2910	2	US-08-466- Sequence 183, Applicat	Sequence 183, Applicat	5.35e+01
45	71	16.0	2910	2	US-08-461- Sequence 183, Applicat	Sequence 183, Applicat	5.35e+01

ALIGNMENTS

RESULT 1
ID US-08-421-144A-8 STANDARD; PRT; 91 AA.
XX
AC xxxxxx
DT
XX
XX
DE Sequence 8, Application US/08421144A
XX
CC Sequence 8, Application US/08421144A
CC Patent No. 5874211
CC GENERAL INFORMATION:
CC APPLICANT: BANDMAN, OLGA
CC APPLICANT: COLEMAN, ROGER
CC TITLE OF INVENTION: NEW CHEMOKINE EXPRESSED IN EOSINOPHILS
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA: US/08/421.144A
CC APPLICATION NUMBER: US/08/421.144A
CC FILING DATE: 13-APR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Luther, Barbara J.
CC REGISTRATION NUMBER: 33954
CC REFERENCE/DOCKET NUMBER: PF-0031 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-852-0195
CC INFORMATION FOR SEQ ID NO: 8:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 91 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein

CC ADDRESSEE: Incyte Pharmaceuticals, Inc.

CC STREET: 3174 Porter Drive

CC CITY: Palo Alto

CC STATE: CA

CC COUNTRY: U.S.

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FastSeq Version 1.5

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/798,143

CC FILING DATE: 10-FEB-1997

CC CLASSIFICATION: 536

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER: 08/347,492

CC FILING DATE: 29-NOV-1994

CC APPLICATION NUMBER: 08/303,241

CC FILING DATE: 07-SEP-1994

CC APPLICATION NUMBER: 08/320,011

CC FILING DATE: 05-OCT-1994

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Luther, Barbara J

CC REGISTRATION NUMBER: 33,954

CC REFERENCE/DOCKET NUMBER: PF-0024

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 415-855-0555

CC TELEFAX: 415-852-0195

CC INFORMATION FOR SEQ ID NO: 12:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 91 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC IMMEDIATE SOURCE:

CC LIBRARY: GENBANK

CC CLONE: GI 134510

CC SEQUENCE 91 AA; 10075 MW; 44825 CN;

CC Query Match 17.3%; Score 77; DB 2; Length 91;

CC Best Local Similarity 38.5%; Pred. No. 1.77e+01;

CC Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

CC Db 19 APASAPSYSDTTPCCFAYIAR-PLP 43

CC Oy 20 SSASPRKFLTNVTCGFATVTRIPLS 45

CC RESULT 7

CC ID US-08-375-346A-5 STANDARD; PRT; 91 AA.

CC XX xxxxxx

CC XX

CC DT

CC XX

CC DE

CC Sequence 5, Application US/08375346A

CC Sequence 5, Application US/08375346A

CC Patent No. 5605817

CC GENERAL INFORMATION:

CC APPLICANT: Coleman, Roger

CC APPLICANT: Wilde, Craig G.

CC APPLICANT: Sellhammer, Jeffrey J.

CC TITLE OF INVENTION: A NEW CHEMOKINE EXPRESSED IN FETAL SPLEEN,

CC TITLE OF INVENTION: ITS PRODUCTION AND USES

CC NUMBER OF SEQUENCES: 9

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

CC STREET: 3330 HILLVIEW AVENUE

CC CITY: PALO ALTO

CC STATE: CALIFORNIA

CC COUNTRY: USA

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FastSeq Version 1.5

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/375,346A

CC FILING DATE: 19-JAN-1995

CC CLASSIFICATION: 435

CC PRIOR APPLICATION DATA:

CC APPLICATION NUMBER:

CC FILING DATE:

CC ATTORNEY/AGENT INFORMATION:

CC NAME: LUTHER, BARBARA J.

CC REGISTRATION NUMBER: 33,954

CC REFERENCE/DOCKET NUMBER: PF-0026 US

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (415) 855-0555

CC TELEFAX: (415) 855-0572

CC TELEX:

CC INFORMATION FOR SEQ ID NO: 5:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 91 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

CC HYPOTHETICAL: NO

CC ANTI-SENSE: NO

CC FRAGMENT TYPE: internal

CC ORIGINAL SOURCE:

CC SEQUENCE 91 AA; 10075 MW; 44825 CN;

CC Query Match 17.3%; Score 77; DB 1; Length 91;

CC Best Local Similarity 38.5%; Pred. No. 1.77e+01;

CC Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

CC Db 19 APASAPSYSDTTPCCFAYIAR-PLP 43

CC Oy 20 SSASPRKFLTNVTCGFATVTRIPLS 45

CC RESULT 8

CC ID US-08-633-682-3 STANDARD; PRT; 91 AA.

CC XX xxxxxx

CC XX

CC DT

CC XX

CC DE

CC XX

CC Sequence 3, Application US/08633682

CC Sequence 3, Application US/08633682

CC Patent No. 5840544

CC GENERAL INFORMATION:

CC APPLICANT: Hawkins, Phillip R.

CC APPLICANT: Bandman, Olga

CC APPLICANT: Murry, Lynn E.

CC TITLE OF INVENTION: NOVEL RANTES HOMOLOG FROM PROSTATE

CC NUMBER OF SEQUENCES: 5

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Incyte Pharmaceuticals, Inc.

CC STREET: 3174 Porter Drive

CC CITY: Palo Alto

CC STATE: CA

CC COUNTRY: U.S.

CC ZIP: 94304

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Diskette

CC COMPUTER: IBM Compatible

CC OPERATING SYSTEM: DOS

CC SOFTWARE: FastSeq Version 1.5

ZIP: 00003
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US92/09487
 FILING DATE: 19921104
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/901,707
 FILING DATE: 19-JUN-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/787,567

CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/646,360
CC	FILING DATE: 13-MAY-1996
CC	CLASSIFICATION: 530
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US94/0534
CC	FILING DATE: 12-MAY-1994
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 08/064,691
CC	FILING DATE: 12-MAY-1993
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 07/988,430
CC	FILING DATE: 09-DEC-1992
CC	PRIOR APPLICATION DATA:
CC	APPLICATION NUMBER: US 07/901,707
CC	FILING DATE: 19-JUN-1992
CC	PRIOR APPLICATION DATA:

NAME: No. 5416202and, Greta E

TELEX: 25-3856


```

##residues      1-326 #label OLA
##cross-references GB:M86261; NID:g162597; PID:g162598
##experimental_source brain
##note          sequence extracted from NCBI backbone (NCBIN:103815,
                  NCBI:P103816)
REFERENCE
#authors      S20390
#journal      Tucker, A.L.; Linden, J.; Robeva, A.S.; D'Angelo, D.D.;
#title        Lynch, K.R.
#abstract     FEBS Lett. (1992) 297:107-111
#description  Cloning and expression of a bovine adenosine A(1) receptor
                  CDNA.
#cross-references MUID:92201360
#accession    S20390
#status      nucleic acid sequence not shown
#molecule_type mRNA
##residues      1-100, 'M', 102-326 #label TUC
##cross-references EMBL:X63592
##experimental_source brain
#CLASSIFICATION #superfamily adenosine receptor A1
#KEYWORDS       adenylylate cyclase inhibitor; G protein-coupled receptor;
                  glycoprotein; transmembrane protein
FEATURE
148,159        #binding_site carbohydrate (Asn) (covalent) #status
                  predicted
SUMMARY
#length 326 #molecular-weight 36579 #checksum 6464
Query Match    19.3%; Score 86; DB 2; Length 326;
Best Local Similarity 31.8%; Pred. No. 9.47e-02;
Matches 14; Conservative 13; Mismatches 14; Indels 3; Gaps 3;
Db 111 IPLAYKTVTPRRAVAITGCWILSFVGLTPMGWNLNAVER 154
QY 14 FPLTFKSSASPRKFLTNVTC-CFA-TVTRIPLSN-KVLTAVD 54

RESULT 3
ENTRY   S56958
TITLE   probable membrane protein YJL175w - yeast (Saccharomyces
        cerevisiae)
ALTERNATE_NAMES hypothetical protein J0502
ORGANISM #formal_name Saccharomyces cerevisiae
DATE      08-Jul-1995 #sequence_revision 08-Sep-1995 #text_change
14-Nov-1997
ACCESSIONS S56958
REFERENCE  S56937
#authors   Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.
#submission submitted to the Protein Sequence Database, September 1995
#accession S56958
#molecule_type DNA
#status    preliminary; translated from GB/EMBL/DBJ
##residues 1-170 #label OBE
##cross-references EMBL:Z49451; NID:g1008372; PID:g1008374; MIPS:YJL175w
GENETICS
#map_position 10L
#KEYWORDS     transmembrane protein
SUMMARY       #length 170 #molecular-weight 17731 #checksum 7245
Query Match    19.1%; Score 85; DB 2; Length 170;
Best Local Similarity 39.3%; Pred. No. 1.36e-01;
Matches 11; Conservative 7; Mismatches 9; Indels 1; Gaps 1;
Db 81 IVGCCFANTVIASLSASRVFGTVASSFR 108
QY 31 VTGCFATVTRIPLS-NKVLTAVDRLR 57

RESULT 4
ENTRY   VCFVER
TITLE   #type complete
CONTAINS env polyprotein - Rous sarcoma virus (strain Prague C)
#KEYWORDS coat protein gp37; coat protein gp85
#formal_name Rous sarcoma virus
ORGANISM #formal_name Rous sarcoma virus
DATE      01-Sep-1981 #sequence_revision 17-Dec-1982 #text_change
04-Oct-1996
ACCESSIONS A03996

```

```

REFERENCE A00632
#authors   Schwartz, D.; Tizard, R.; Gilbert, W.
#submission submitted to the Nucleic Acid Sequence Database, September
1982
#accession A03996
#molecule_type genomic RNA
##residues      1-603 #label SCH
##note          as a result of base variations, a different version of
                  this sequence may exist having 17-Glu, 134-Ser,
                  158-Ser, 334-Thr, 383-Thr, 392-Val, 522-Glu, 541-Leu,
                  and 567-Val
GENETICS
#gene        env
#CLASSIFICATION #superfamily type C retrovirus env polyprotein
#KEYWORDS     polyprotein
FEATURE
65-603        #product coat protein gp85 #status predicted #label P85\
406-603        #product coat protein gp37 #status predicted #label P37
SUMMARY       #length 603 #molecular-weight 65660 #checksum 5554
Query Match    18.9%; Score 84; DB 1; Length 603;
Best Local Similarity 35.5%; Pred. No. 1.96e-01;
Matches 11; Conservative 10; Mismatches 9; Indels 1; Gaps 1;
Db 304 PFSTANTGNS-IGNVSGCGCEPITILPLG 333
QY 15 PLTFKSSASPRKFLTNVTCGCFATVTRIPLS 45

RESULT 5
ENTRY   T00094
TITLE   #type complete
#description  endostyle-specific protein - sea squirt (Cliona intestinalis)
#formal_name Cliona intestinalis
#submission submitted to the EMBL Data Library, February 1998
#accession T00094
#status    preliminary; translated from GB/EMBL/DBJ
##residues 1-650 #label OGA
##cross-references EMBL:AB010895; NID:d1179651; PID:d1025751
##experimental_source endostyle
SUMMARY       #length 650 #molecular-weight 75483 #checksum 3537
Query Match    18.9%; Score 84; DB 2; Length 650;
Best Local Similarity 31.6%; Pred. No. 1.96e-01;
Matches 12; Conservative 10; Mismatches 12; Indels 4; Gaps 4;
Db 344 YKTRYCFVRYCLOPFKFCFNPKY-TGLKTCVFPVAVR 380
QY 7 FQSRFC-I-FPL-TFKSSASPRKFLTNVTCGCFATVTR 41

RESULT 6
ENTRY   S70364
TITLE   #type complete
ALTERNATE_NAMES cannabinoind receptor CB2, peripheral - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE      21-Apr-1997 #sequence_revision 09-May-1997 #text_change
17-Mar-1999
ACCESSIONS S70364; S54163
REFERENCE  S70364
#authors   Shire, D.; Calandra, B.; Rinaldi-Carmona, M.; Oustric, D.;
                  Pessegue, B.; Bonnin-Cabanne, O.; le Fur, G.; Caput, D.;
                  Ferrara, P.
#journal     Biochim. Biophys. Acta (1996) 1307:132-136
#title       Molecular cloning, expression and function of the murine CB2
                  peripheral cannabinoind receptor.

```

[illegible]

authors Lee-Huang, S.; Huang, P.L.; Chen, H.C.; Huang, P.L.;
 Bourinbaiaf, A.; Huang, H.I.; Kung, H.F.
#journal Gene (1995) 161:151-156
#title Anti-HIV and anti-tumor activities of recombinant MAP30 from bitter melon.
#cross-references MUID:95394347
#accession JC4235
##molecule_type DNA
##residues 1-286 #label LEE
##experimental_source mature seeds
##note The authors translated the codon TAC for residue 37 as Thr and ACT for residue 188 as Ile
REFERENCE
#authors Lee-Huang, S.; Huang, P.L.; Nara, P.L.; Chen, H.C.; Kung, H.; Huang, P.; Huang, H.I.; Huang, P.L.
#journal FEBS Lett. (1990) 272:12-18
#title MAP 30: a new inhibitor of HIV-1 infection and replication.
#cross-references MUID:91032105
#accession S12869
##molecule_type protein
##residues 24-36, 'n', 38-66, 'p' #label LEW
COMMENT This plant protein has anti-HIV activity. It possesses antiviral action, anti-tumor activity, topological inactivation of viral DNA, inhibition of viral integrase and cell-free ribosome-inactivation activities. It is capable of acting against multiple stages of the viral life cycle, on acute infection as well as replication in chronically infected cells.
COMMENT This protein has conserved unique residues Trp-213 and Met-277.
GENETICS
#gene map30
CLASSIFICATION #superfamily rRNA N-glycosidase; rRNA N-glycosidase homology
KEYWORDS antiviral; glycoprotein; glycosidase; hydrolase
FEATURE
 1-23 #domain signal sequence #status predicted #label SIG\
 24-286 #product rRNA N-glycosidase (EC 3.2.2.22) map30 #status predicted #label MAT\
 27-264 #domain rRNA N-glycosidase homology #label RNG\
 74 #binding_site carbohydrate (Asn) (covalent) #status predicted\
 93,181,184 #active_site Tyr, Glu, Arg #status predicted
SUMMARY #length 286 #molecular-weight 32018 #checksum 4956
Query Match 17.8%; Score 79; DB 2; Length 286;
Best Local Similarity 28.9%; Pred. No. 1.15e+00;
Matches 11; Conservative 12; Mismatches 14; Indels 1; Gaps 1;
Db 47 TLPFSKVVDPILLYTSPISRRFILLNLTXYAYETIS 84
QY 4 TLVFSQRFICFLPTEKSSASPRKF-LTNVTGCCFATVT 40
RESULT 11
ENTRY C70108 #type complete
TITLE peptidase homolog - Lyme disease spirochete
ORGANISM #formal_name Borrelia burgdorferi #common_name Lyme disease spirochete
DATE 13-Feb-1998 #sequence_revision 13-Feb-1998 #text_change 05-Jun-1998
ACCESSIONS C70108
REFERENCE A70100
#authors Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; White, O.; Ketchum, K.A.; Dodson, R.; Hickey, E.K.; Winn, M.; Dougherty, B.; Tomb, J.F.; Fleischmann, R.D.; Richardson, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vugt, R.V.; Palmer, M.; Adams, M.D.; Gocayne, J.; Weidman, J.; Utterback, T.; Watthey, L.; McDonald, L.; Artiach, P.; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.; Smith, H.O.; Venter, J.C.
#journal Nature (1997) 390:580-586
#title Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi.
#cross-references MUID:98065943

#accession C70108
##status preliminary; nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-592 #label KLE
##cross-references GB:AE001119; GB:AE000783; NID:92687936; PID:92687937; TIGR:BB0067
##experimental_source strain B31
SUMMARY #length 592 #molecular-weight 67623 #checksum 1324
Query Match 17.8%; Score 79; DB 2; Length 592;
Best Local Similarity 46.9%; Pred. No. 1.15e+00;
Matches 15; Conservative 5; Mismatches 10; Indels 2; Gaps 2;
Db 40 STRKFTFGSGS-FGTVI-ITLSKAVLTDCGR 69
QY 23 SPREFLNVTCGFATVTRPLSNKVLTAVD 54
RESULT 12
ENTRY W2BEB6 #type complete
TITLE 367K tegument protein - equine herpesvirus 1 (strain Ab4p)
ORGANISM #formal_name equine herpesvirus 1
#note host Equus caballus (domestic horse)
DATE 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 05-Sep-1997
ACCESSIONS G36797
REFERENCE A36805
#authors Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
#submission submitted to GenBank, March 1992
#description The DNA sequence of equine herpesvirus-1.
#accession G36797
##molecule_type DNA
##residues 1-3421 #label TEL
##cross-references GB:M86664; NID:9330791; PID:9330816
REFERENCE A41831
#authors Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
#journal Virology (1992) 189:304-316
#title The DNA sequence of equine herpesvirus-1.
#cross-references MUID:9295566
#contents annotation; possible protein-coding frames
#note neither amino acid nor nucleotide sequence is given
GENETICS 24
#gene 24
CLASSIFICATION #superfamily varicella-zoster virus gene 22 protein
SUMMARY #length 3421 #molecular-weight 367079 #checksum 2318
Query Match 17.8%; Score 79; DB 1; Length 3421;
Best Local Similarity 30.8%; Pred. No. 1.15e+00;
Matches 12; Conservative 14; Mismatches 11; Indels 2; Gaps 2;
Db 3373 FRSTLSAMALL-IAACR-TIVRRLRATRRVLTINRSL 3409
QY 18 FKSSAPRKFLTNVTCGFATVTRPLSNKVLTAVD 56
RESULT 13
ENTRY S58343 #type complete
TITLE Shiga-like toxin II chain A - Enterobacter cloacae
ORGANISM #formal_name Enterobacter cloacae
DATE 14-Jan-1996 #sequence_revision 01-Mar-1996 #text_change 08-Sep-1997
ACCESSIONS S58343
REFERENCE S58343
#authors Paton, A.W.; Paton, J.C.
#submission submitted to the EMBL Data Library, August 1995
#description Enterobacter cloacae producing a Shiga-like toxin II-related cytotoxin associated with a case of hemolytic uremic syndrome.
#accession S58343
##status preliminary
##molecule_type DNA
##residues 1-319 #label PAT

WISORLH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:20:11 1999; MasPar time 4.57 Seconds
364.887 Million cell updates/sec
Tabular output not generated.

Title: >US-09-209-961-7
Description: (1-59) from US09209961.pep
Perfect Score: 445
Sequence: 1 MKCTLVQSRFCIFPLFKS.....TRIPLSNKVLTAVDRLRCP 59

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 36.740; Variance 52.488; scale 0.700

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	89	20.0	111	1	YZPU_ECOLI	VERY HYPOTHETICAL 13.2	4.07e-03
2	86	19.3	326	1	AAIR_BOVIN	ADENOSINE A1 RECEPTOR	1.46e-02
3	85	19.1	170	1	YJRS_YEAST	HYPOTHETICAL 17.7 KD P	2.21e-02
4	84	18.9	603	1	ENV_RSVF	ENV POLYPROTEIN [CONTA	3.35e-02
5	83	18.7	347	1	CB2R_MOUSE	CANNABINOID RECEPTOR 2	5.06e-02
6	79	17.8	3421	1	TEGU_HSVB	LARGE TEGUMENT PROTEIN 2	5.46e-01
7	78	17.5	322	1	GDC_RAT	GRAVE'S DISEASE CARRIE	3.77e-01
8	78	17.5	332	1	GDC_HUMAN	GRAVE'S DISEASE CARRIE	3.77e-01
9	78	17.5	403	1	ZXDB_HUMAN	ZINC FINGER X-LINKED P	3.77e-01
10	78	17.5	457	1	ZXDA_HUMAN	ZINC FINGER X-LINKED P	3.77e-01
11	77	17.3	91	1	SISD_HUMAN	T-CELL SPECIFIC RANTES	5.57e-01
12	77	17.3	286	1	RIP2_MOMBA	RIBOSOME-INACTIVATING	5.57e-01
13	77	17.3	330	1	GDC_BOVIN	GRAVE'S DISEASE CARRIE	5.57e-01
14	77	17.3	606	1	NUSM_RHUN	NADH-UBIQUINONE OXIDOR	5.57e-01
15	76	17.1	326	1	AAIR_HUMAN	ADENOSINE A1 RECEPTOR	8.20e-01
16	76	17.1	400	1	HOFC_ECOLI	PROTEIN TRANSPORT PROT	8.20e-01
17	76	17.1	684	1	TGLA_HUMAN	PROTEIN-GLUTAMINE GLUT	8.20e-01
18	75	16.9	319	1	SLTA_BP933	SHIGA-LIKE TOXIN II SU	1.20e+00
19	75	16.9	912	1	KPCM_HUMAN	PROTEIN KINASE C, MU T	1.20e+00
20	75	16.9	1887	1	YDHA_SCHPO	HYPOTHETICAL 192.5 KD	1.20e+00
21	74	16.6	308	1	YL17_HELPY	HYPOTHETICAL PROTEIN H	1.76e+00
22	74	16.6	531	1	UL95_HCMVA	PROTEIN UL95	1.76e+00
23	74	16.6	639	1	YAJ9_SCHPO	HYPOTHETICAL 74.4 KD P	1.76e+00

24	74	16.6	942	1	ISP2_HUMAN	TYPE II INOSITOL-1,4,5	1.76e+00
25	73	16.4	1032	1	MT18_YEAST	DNA REPAIR/TRANSCRIPTI	2.56e+00
26	72	16.2	64	1	DROS_DROME	DROSOCIN PRECURSOR	3.70e+00
27	72	16.2	121	1	YAF1_CAEEL	HYPOTHETICAL 13.4 KD P	3.70e+00
28	72	16.2	312	1	UCP3_HUMAN	MITOCHONDRIAL UNCOUPLI	3.70e+00
29	72	16.2	345	1	LEU3_LACLA	3-ISOPROPYLMALATE DEHY	3.70e+00
30	71	16.0	91	1	SISD_CAVPO	T-CELL SPECIFIC RANTES	5.34e+00
31	71	16.0	115	1	KV2A_HUMAN	IG KAPPA CHAIN V-II RE	5.34e+00
32	71	16.0	282	1	SRG2_CAEEL	SRG-2 PROTEIN	5.34e+00
33	71	16.0	326	1	AAIR_CANFA	ADENOSINE A1 RECEPTOR	5.34e+00
34	71	16.0	326	1	AAIR_CANFA	ADENOSINE A1 RECEPTOR	5.34e+00
35	71	16.0	328	1	AAIR_RABIT	ADENOSINE A1 RECEPTOR	5.34e+00
36	71	16.0	360	1	CB2R_HUMAN	CANNABINOID RECEPTOR 2	5.34e+00
37	71	16.0	529	1	YABD_SCHPO	HYPOTHETICAL 58.6 KD P	5.34e+00
38	71	16.0	592	1	PUR2_YEAST	PHOSPHORIBOSYLAMINOIM	5.34e+00
39	71	16.0	1601	1	RRPO_TVCV	PUTATIVE RNA-DIRECTED	5.34e+00
40	71	16.0	1636	1	YN37_YEAST	HYPOTHETICAL 186.8 KD	5.34e+00
41	71	16.0	4466	1	DYHC_TRIGR	DYNEIN BETA CHAIN, CIL	5.34e+00
42	71	16.0	4466	1	DYHC_ANTCR	DYNEIN BETA CHAIN, CIL	5.34e+00
43	70	15.7	128	1	Y070_TREPA	HYPOTHETICAL PROTEIN T	7.67e+00
44	70	15.7	297	1	VRPR_SALDU	VIRULENCE GENES TRANSC	7.67e+00
45	70	15.7	444	1	AFLR_ASPPA	AFLATOXIN BIOSYNTHESIS	7.67e+00

ALIGNMENTS

RESULT 1
ID YZPU_ECOLI STANDARD; PRT; 111 AA.
AC P24254;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE VERY HYPOTHETICAL 13.2 KD PROTEIN IN PUTA-PUTP INTERGENIC REGION.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
OC ESCHERICHIA.
RP SEQUENCE FROM N.A.
RN [1]
RC STRAIN=K12;
RX MEDLINE; 88142554
RX NAKAO T., YAMATO I., ANRAKU Y.;
RT "Nucleotide sequence of putc, the regulatory region for the put
regulation of Escherichia coli K12."
RU MOL. GEN. GENET. 210:364-368(1987).
CC -!- CAUTION: THIS SEQUENCE, ACCORDING TO THE EMBL DATABASE (K. RUDD)
IS PROBABLY NOT A REAL PROTEIN; THEREFORE THIS ENTRY WILL PROBABLY
BE DELETED IN FUTURE RELEASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; M35174; G147431;
DR EMBL; X05553; G42601;
DR FIR; S06385; S06385;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 111 AA; 13203 MW; 6E3E921D CRC32;

Query Match 20.08; Score 89; DB 1; Length 111;
Best Local Similarity 35.34; Pred. No. 4.07e-03;
Matches 18; Conservative 17; Mismatches 11; Indels 5; Gaps 4;
DB 12 ALVTHNRENT-PFTLNFSF-ORHYFS--SGCTLSHFLRLHLS-KMLTAAEK 57
OY 4 TLVFQSRFCIFPLFKSASPRFLTNVTCGFATVTRPLSNKVLTAVD 54
RESULT 2
ID AAIR_BOVIN STANDARD; PRT; 326 AA.

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; V01197; G61703; -
 DR PIR; A03996; VCFVER; -
 DR HSSP; P03385; IMOF; -
 KW COAT PROTEIN; POLYPROTEIN; TRANSMEMBRANE; SIGNAL.
 FT SIGNAL 1 64
 FT CHAIN 65 603 SURFACE PROTEIN GP85.
 FT CHAIN 406 603 MEMBRANE PROTEIN GP37.
 FT TRANSMEM 557 577 POTENTIAL.
 SQ SEQUENCE 603 AA: 65660 MW: C80D2F1A CRC32;

Query Match 18.9%; Score 84; DB 1; Length 603;
 Best Local Similarity 35.5%; Pred. No. 3.35e-02;
 Matches 11; Conservative 10; Mismatches 9; Indels 1; Gaps 1;

Db 304 PFSFTANTGSN-LGNVSCCGCEPITILPLG 333
 QY 15 PLTFKSSASPRKFLNTVTCGFATVTRIPLS 45

RESULT 5
 ID CB2R_MOUSE STANDARD; PRT; 347 AA.
 AC P47936;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE CANNABINOID RECEPTOR 2 (CB2) (CB-2).
 GN CNR2.

OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6; TISSUE-SPLEEN;
 RX MEDLINE; 96283804.
 RA SHIRE D., CALANDRA B., RINALDI-CARMONA M., OUSTRIC D., PESSEGUE B.,
 RA CABANNE O., LE FUR G., CAPUT D., FERRARA P.;
 RT "Molecular cloning, expression and function of the murine CB2
 RT peripheral cannabinoid receptor.";
 RL BIOCHIM. BIOPHYS. ACTA 1307:132-136(1996).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN-129/SV;
 RA BUCKLEY N.E., BONNER T.I.;
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-NFS/N;
 RX MEDLINE; 97404692.
 RA VALK P.J.M., HOL S., VANKAN Y., IHLE J.N., ASKEW D., JENKINS N.A.,
 RA GILBERT D.J., COPELAND N.G., DE BOTH N.J., LOWENBERG B., DELWEL R.;
 RT "The genes encoding the peripheral cannabinoid receptor and alpha-L-
 RT fucosidase are located near a newly identified common virus
 RT integration site, Evill.";
 RL J. VIROL. 71:6796-6804(1997).
 CC [1] FUNCTION: INVOLVED IN CANNABINOID-INDUCED CNS EFFECTS. COULD BE A
 CC RECEPTOR FOR ANANDAMIDE.

CC [1] SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC [1] SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; X86405; G791082; -
 DR EMBL; U21681; G710428; -

DR EMBL; X93168; E213356; -
 DR GCRDB; GCR_1704; -
 DR GCRDB; GCR_1725; -
 DR GCRDB; GCR_2451; -
 DR MGD; MGI:104650; CNR2.
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm1; 1.

KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN.
 FT DOMAIN 1 33
 FT TRANSMEM 34 59
 FT DOMAIN 60 71
 FT TRANSMEM 72 92
 FT DOMAIN 93 104
 FT TRANSMEM 105 129
 FT DOMAIN 130 149
 FT TRANSMEM 150 172
 FT DOMAIN 173 188
 FT TRANSMEM 189 214
 FT DOMAIN 215 246
 FT TRANSMEM 247 267
 FT DOMAIN 268 279
 FT TRANSMEM 280 301
 FT DOMAIN 302 347
 FT CARBOXYD 11 11
 FT CONFLICT 54 64
 FT CONFLICT 126 126
 FT CONFLICT 272 272
 FT CONFLICT 332 332
 SQ SEQUENCE 347 AA: 38213 MW: 753AD117 CRC32;

Query Match 18.7%; Score 83; DB 1; Length 347;
 Best Local Similarity 34.5%; Pred. No. 5.06e-02;
 Matches 19; Conservative 12; Mismatches 20; Indels 4; Gaps 4;

Db 84 SVIFACNFVIFHV-FHGVDNSAIFILKI-GSVTMTFT-ASVGSLLLTAVDRYL-C 134
 QY 4 TLVFCRSFCIFPLTFKSSASPRKFLNTVTCGFATVTRIPLSKVLTAVDRLSRC 58

RESULT 6
 ID TEGU_HVHB STANDARD; PRT; 3421 AA.
 AC F28955;
 DT 01-DEC-1992 (REL. 24, CREATED)
 DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE LARGE TEGUMENT PROTEIN.
 GN 24.
 OS EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
 OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.

[1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92295566.
 RA TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAVIDSON A.J.;
 RT "The DNA sequence of equine herpesvirus-1.";
 RL VIROLOGY 189:304-316(1992).

CC [1] FUNCTION: TEGUMENT PROTEIN.
 CC [1] SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,
 CC EHV-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)

DR EMBL; M86564; G330816; -
 DR PIR; G36797; WZBB66.
 DR HSSP; P04002; 1WFA.
 SQ SEQUENCE 3421 AA: 367078 MW: 9F15166B CRC32;

Query Match 17.88; Score 79; DB 1; Length 3421;
Best Local Similarity 30.88; Pred. No. 2:54e-01;
Matches 12; Conservative 14; Mismatches 11; Indels 2; Gaps 2;

DB 3373 FRSTLSAMALL-IAACR-TIVRLRATRLVLTIDNRS 3409
QY 18 FKASAPRFLTNVTCGCFATVTRIPLSNKLTVADRS 56

RESULT 7
ID GDC_RAT STANDARD; PRT; 322 AA.
AC P16261;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GRAVE'S DISEASE CARRIER PROTEIN (GDC) (MITOCHONDRIAL SOLUTE CARRIER
DE PROTEIN HOMOLOG) (FRAGMENT).
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RX MEDLINE; 90114217.
RA ZARRILLI R., OATES E.L., MCBRIDE O.W., LERMAN M.I., CRAN J.Y.,
RA SANTISTEBAN P., URSINI M.V., NOTKINS A.L., KOHN L.D.;
RT "Sequence and chromosomal assignment of a novel cDNA identified by
RT immunoscreening of a thyroid expression library: similarity to a
RT family of mitochondrial solute carrier proteins.";
RL MOL. ENDOCRINOL. 3:1498-1508(1989).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (POTENTIAL).
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M32973; G205529;
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
DR PFAM; PF00153; mito_carr; 1
KW MITOCHONDRION; TRANSMEMBRANE; TRANSPORT; REPEAT.
FT REPEAT 1 123
FT REPEAT 124 222
FT REPEAT 223 322
FT NON_TER 322 322
SQ SEQUENCE 322 AA; 35056 MW; 346AFDA6 CRC32;

Query Match 17.58; Score 78; DB 1; Length 322;
Best Local Similarity 48.18; Pred. No. 3.77e-01;
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

DB 38 RSFLAGSIAGCC-AKTTVAPLDRVKVL 63
QY 25 RKFLT-NVTGCCFATVTRIPLSN-KVL 49

RESULT 8
ID GDC_HUMAN STANDARD; PRT; 332 AA.
AC P16260;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
DE GRAVE'S DISEASE CARRIER PROTEIN (GDC) (MITOCHONDRIAL SOLUTE CARRIER
DE PROTEIN HOMOLOG).
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94093547.
RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of

OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-THYROID;
RX MEDLINE; 90114217.
RA ZARRILLI R., OATES E.L., MCBRIDE O.W., LERMAN M.I., CHAN J.Y.,
RA SANTISTEBAN P., URSINI M.V., NOTKINS A.L., KOHN L.D.;
RT "Sequence and chromosomal assignment of a novel cDNA identified by
RT immunoscreening of a thyroid expression library: similarity to a
RT family of mitochondrial solute carrier proteins.";
RL MOL. ENDOCRINOL. 3:1498-1508(1989).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. MITOCHONDRIAL
CC INNER MEMBRANE (POTENTIAL).
CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
CC -1- DISEASE: POSSIBLE ROLE IN GRAVES' DISEASE.
CC -1- SIMILARITY: BELONGS TO THE MITOCHONDRIAL CARRIER FAMILY.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN FROM POSITION 320
CC ONWARD DUE TO A FRAMESHIFT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M31659; G386960; ALT_FRAME.
DR PIR; A40141; A40141.
DR MIM; L39080;
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
DR PFAM; PF00153; mito_carr; 1
KW MITOCHONDRION; TRANSMEMBRANE; TRANSPORT; REPEAT.
FT REPEAT 1 123
FT REPEAT 124 219
FT REPEAT 220 332
SQ SEQUENCE 332 AA; 36235 MW; 7E1DB8DC CRC32;

Query Match 17.58; Score 78; DB 1; Length 332;
Best Local Similarity 48.18; Pred. No. 3.77e-01;
Matches 13; Conservative 4; Mismatches 7; Indels 3; Gaps 3;

DB 38 RSFLAGSIAGCC-AKTTVAPLDRVKVL 63
QY 25 RKFLT-NVTGCCFATVTRIPLSN-KVL 49

RESULT 9
ID ZYDB_HUMAN STANDARD; PRT; 403 AA.
AC P98169;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ZINC FINGER X-LINKED PROTEIN ZYDB (FRAGMENT).
GN ZYDB.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE; 94093547.
RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of

the human X chromosome: isolation, characterization and X-inactivation studies. HUM. MOL. GENET. 2:1611-1618(1993).
SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
TISSUE SPECIFICITY: UBIQUITOUS.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL: L14788; G292933;
PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
PFAM; PF00096; zf-C2H2; 9.
HSP; P08047; LSP2.
ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
NON_TER 1
DOMAIN <1 266 ZINC-FINGERS.
ZNFING <1 22 C2H2-TYPE.
ZNFING 28 52 C2H2-TYPE.
ZNFING 58 80 C2H2-TYPE.
ZNFING 87 111 C2H2-TYPE.
ZNFING 118 142 C2H2-TYPE.
ZNFING 148 172 C2H2-TYPE.
ZNFING 178 202 C2H2-TYPE.
ZNFING 208 232 C2H2-TYPE.
ZNFING 241 266 C2H2-TYPE.
SEQUENCE 403 AA; 44403 MW; 57ADG326 CRC32;
Query Match 17.5%; Score 78; DB 1; Length 403;
Best Local Similarity 25.4%; Pred. No. 3.77e-01;
Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;
Db 125 CSKQYD-KACRLKHLRSHTGEPFLCDGCGWNTSMKSL-LRHKRHHDDRRMCP 181
QY 3 CTLVFSQRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59
RESULT 10
ID ZNDA_HUMAN STANDARD; PRT; 457 AA.
AC P98168;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ZINC FINGER X-LINKED PROTEIN ZNDA (FRAGMENT).
GN ZNDA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
PP SEQUENCE FROM N.A.
RX MEDLINE; 94093547.
RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of the human X chromosome: isolation, characterization and X-inactivation studies."
RL HUM. MOL. GENET. 2:1611-1618(1993).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL: L14787; G292931;
PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
PFAM; PF00096; zf-C2H2; 9.
HSP; P08047; LSP2.
ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
NON_TER 1
DOMAIN <1 266 ZINC-FINGERS.
ZNFING <1 22 C2H2-TYPE.
ZNFING 28 52 C2H2-TYPE.
ZNFING 58 80 C2H2-TYPE.
ZNFING 87 111 C2H2-TYPE.
ZNFING 118 142 C2H2-TYPE.
ZNFING 148 172 C2H2-TYPE.
ZNFING 178 202 C2H2-TYPE.
ZNFING 208 232 C2H2-TYPE.
ZNFING 241 266 C2H2-TYPE.
SEQUENCE 457 AA; 49978 MW; 10CC474E CRC32;
Query Match 17.5%; Score 78; DB 1; Length 457;
Best Local Similarity 25.4%; Pred. No. 3.77e-01;
Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;
Db 125 CSKQYD-KACRLKHLRSHTGEPFLCDGCGWNTSMKSL-LRHKRHHDDRRMCP 181
QY 3 CTLVFSQRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59
RESULT 11
ID SISD_HUMAN STANDARD; PRT; 91 AA.
AC P13501;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL SPECIFIC RANTES PROTEIN PRECURSOR (SIS-DELTA) (SMALL INDUCIBLE CYTOKINE A5).
GN SCYA5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
PP SEQUENCE FROM N.A.
RX MEDLINE; 88285659.
RA SCHALL T.J., JONGSTRA J., DYER B.J., JORGENSEN J., CLAYBERGER C., DAVIS M.M., KRENSKY A.M.;
RT "A human T cell-specific molecule is a member of a new gene family."
RL J. IMMUNOL. 141:1018-1025(1988).
RN [2]
RF STRUCTURE BY NMR.
RA MEDLINE; 95352612.
RT "The three-dimensional solution structure of RANTES."
RL BIOCHEMISTRY 34:9307-9314(1995).
RN [3]
PP STRUCTURE BY NMR.
RX MEDLINE; 95244456.
RA SKELTON N.J., ASPIRAS F., OGEE J., SCHALL T.J.;
RT "Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type."
RL BIOCHEMISTRY 34:5329-5342(1995).
CC -!- FUNCTION: CHEMOATTRACTANT FOR BLOOD MONOCYTES, MEMORY T HELPER CELLS AND EOSINOPHILS. CAUSES THE RELEASE OF HISTAMINE FROM BASOPHILS AND ACTIVATES EOSINOPHILS.
CC -!- TISSUE SPECIFICITY: T-CELL AND MACROPHAGE SPECIFIC.
CC -!- INDUCTION: BY MITOGENS.
CC -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE C-C) (CHEMOKINE CC).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

the human X chromosome: isolation, characterization and X-inactivation studies. HUM. MOL. GENET. 2:1611-1618(1993).
SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
TISSUE SPECIFICITY: UBIQUITOUS.
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL: L14788; G292933;
PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
PFAM; PF00096; zf-C2H2; 9.
HSP; P08047; LSP2.
ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
NON_TER 1
DOMAIN <1 266 ZINC-FINGERS.
ZNFING <1 22 C2H2-TYPE.
ZNFING 28 52 C2H2-TYPE.
ZNFING 58 80 C2H2-TYPE.
ZNFING 87 111 C2H2-TYPE.
ZNFING 118 142 C2H2-TYPE.
ZNFING 148 172 C2H2-TYPE.
ZNFING 178 202 C2H2-TYPE.
ZNFING 208 232 C2H2-TYPE.
ZNFING 241 266 C2H2-TYPE.
SEQUENCE 403 AA; 44403 MW; 57ADG326 CRC32;
Query Match 17.5%; Score 78; DB 1; Length 403;
Best Local Similarity 25.4%; Pred. No. 3.77e-01;
Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;
Db 125 CSKQYD-KACRLKHLRSHTGEPFLCDGCGWNTSMKSL-LRHKRHHDDRRMCP 181
QY 3 CTLVFSQRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59
RESULT 10
ID ZNDA_HUMAN STANDARD; PRT; 457 AA.
AC P98168;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE ZINC FINGER X-LINKED PROTEIN ZNDA (FRAGMENT).
GN ZNDA.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
PP SEQUENCE FROM N.A.
RX MEDLINE; 94093547.
RA GREIG G.M., SHARP C.B., CARREL L., WILLARD H.F.;
RT "Duplicated zinc finger protein genes on the proximal short arm of the human X chromosome: isolation, characterization and X-inactivation studies."
RL HUM. MOL. GENET. 2:1611-1618(1993).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
EMBL: L14787; G292931;
PROSITE; PS00028; ZINC_FINGER_C2H2; 8.
PFAM; PF00096; zf-C2H2; 9.
HSP; P08047; LSP2.
ZINC-FINGER; METAL-BINDING; DNA-BINDING; REPEAT; NUCLEAR PROTEIN.
NON_TER 1
DOMAIN <1 266 ZINC-FINGERS.
ZNFING <1 22 C2H2-TYPE.
ZNFING 28 52 C2H2-TYPE.
ZNFING 58 80 C2H2-TYPE.
ZNFING 87 111 C2H2-TYPE.
ZNFING 118 142 C2H2-TYPE.
ZNFING 148 172 C2H2-TYPE.
ZNFING 178 202 C2H2-TYPE.
ZNFING 208 232 C2H2-TYPE.
ZNFING 241 266 C2H2-TYPE.
SEQUENCE 457 AA; 49978 MW; 10CC474E CRC32;
Query Match 17.5%; Score 78; DB 1; Length 457;
Best Local Similarity 25.4%; Pred. No. 3.77e-01;
Matches 15; Conservative 19; Mismatches 21; Indels 4; Gaps 3;
Db 125 CSKQYD-KACRLKHLRSHTGEPFLCDGCGWNTSMKSL-LRHKRHHDDRRMCP 181
QY 3 CTLVFSQRCIFPLTFKSSASPRKFLNTVGTCC--FATVTRPLSNKVLTAVDRLRCP 59
RESULT 11
ID SISD_HUMAN STANDARD; PRT; 91 AA.
AC P13501;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE T-CELL SPECIFIC RANTES PROTEIN PRECURSOR (SIS-DELTA) (SMALL INDUCIBLE CYTOKINE A5).
GN SCYA5.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
[1]
PP SEQUENCE FROM N.A.
RX MEDLINE; 88285659.
RA SCHALL T.J., JONGSTRA J., DYER B.J., JORGENSEN J., CLAYBERGER C., DAVIS M.M., KRENSKY A.M.;
RT "A human T cell-specific molecule is a member of a new gene family."
RL J. IMMUNOL. 141:1018-1025(1988).
RN [2]
RF STRUCTURE BY NMR.
RA MEDLINE; 95352612.
RT "The three-dimensional solution structure of RANTES."
RL BIOCHEMISTRY 34:9307-9314(1995).
RN [3]
PP STRUCTURE BY NMR.
RX MEDLINE; 95244456.
RA SKELTON N.J., ASPIRAS F., OGEE J., SCHALL T.J.;
RT "Proton NMR assignments and solution conformation of RANTES, a chemokine of the C-C type."
RL BIOCHEMISTRY 34:5329-5342(1995).
CC -!- FUNCTION: CHEMOATTRACTANT FOR BLOOD MONOCYTES, MEMORY T HELPER CELLS AND EOSINOPHILS. CAUSES THE RELEASE OF HISTAMINE FROM BASOPHILS AND ACTIVATES EOSINOPHILS.
CC -!- TISSUE SPECIFICITY: T-CELL AND MACROPHAGE SPECIFIC.
CC -!- INDUCTION: BY MITOGENS.
CC -!- SIMILARITY: BELONGS TO THE INTERCRINE BETA FAMILY (SMALL CYTOKINE C-C) (CHEMOKINE CC).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
or send an email to license@sib-sib.ch).
-----
CC EMBL; M21121; G339421;
DR PIR; A28815; A28815;
DR PDB; 1HRJ; 14-OCT-96.
DR PDB; 1HRJ; 03-JUN-95.
DR PDB; 1RTO; 03-JUN-95.
DR MIM; 187011;
DR PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR PFAM; PF00048; i18; 1.
KW CYTOKINE: CHEMOTAXIS; T-CELL; SIGNAL; INFLAMMATORY RESPONSE;
KW 3D-STRUCTURE. 1. 23
FT SIGNAL 24 91 T-CELL SPECIFIC RANTES PROTEIN.
FT CHAIN 33 57
FT DISULFID 34 73
FT DISULFID 34 73
SQ SEQUENCE 91 AA; 10075 MW; 6BDFE23E CRC32;

Query Match 17.3%; Score 77; DB 1; Length 91;
Best Local Similarity 38.5%; Pred. No. 5.57e-01;
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 APASAPYSSDTPCCFAYIAR-PLP 43
QY 20 SSAPRKLNTVNGCCFATVTRPLS 45

RESULT 12
ID RIP2_MOMBA STANDARD; PRT; 286 AA.
AC F29339;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RIBOSOME-INACTIVATING PROTEIN MOMORDIN II PRECURSOR (RRNA
DE N-GLYCOSIDASE) (EC 3.2.2.22).
OS MOMORDICA BALSAMINA (BITTER GOURD) (BALSAM PEAR).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
CN CUCURBITALES; CUCURBITACEAE; MOMORDICA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RX MEDLINE; 93027170.
RA "Momordin II, a ribosome inactivating protein from Momordica
RT balsamina, is homologous to other plant proteins.";
RL NUCLEIC ACIDS RES. 20:4662-4662(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RRNA.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; X66035; G387;
DR PIR; S26596; S26596.
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
DR PFAM; PF00153; mito_carr; 1.
RW MITOCHONDRIUM; TRANSMEMBRANE; TRANSPORT; REPEAT.
FT REPEAT 122 217
FT REPEAT 121 121
FT REPEAT 218 330
SQ SEQUENCE 330 AA; 36085 MW; 428468D4 CRC32;

Query Match 17.3%; Score 77; DB 1; Length 330;
Best Local Similarity 48.1%; Pred. No. 5.57e-01;
Matches 13; Conservative 3; Mismatches 8; Indels 3; Gaps 3;

Db 36 RSFLAGGIAGCC-AKTTPVAPLDVRKVL 61
QY 25 RKLTN-VTGCCFATVTRPLSN-KVL 49

RESULT 14
ID N55M_RHUN STANDARD; PRT; 606 AA.
AC Q96069;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN MTND5 OR NDS OR NADH5.
OS RHINOCEROS UNICORNIS (GREATER INDIAN RHINOCEROS).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;

or send an email to license@sib-sib.ch).
-----
CC EMBL; M21121; G339421;
DR PIR; A28815; A28815;
DR PDB; 1HRJ; 14-OCT-96.
DR PDB; 1HRJ; 03-JUN-95.
DR PDB; 1RTO; 03-JUN-95.
DR MIM; 187011;
DR PROSITE; PS00472; SMALL_CYTOKINES_CC; 1.
DR PFAM; PF00048; i18; 1.
KW CYTOKINE: CHEMOTAXIS; T-CELL; SIGNAL; INFLAMMATORY RESPONSE;
KW 3D-STRUCTURE. 1. 23
FT SIGNAL 24 91 T-CELL SPECIFIC RANTES PROTEIN.
FT CHAIN 33 57
FT DISULFID 34 73
FT DISULFID 34 73
SQ SEQUENCE 91 AA; 10075 MW; 6BDFE23E CRC32;

Query Match 17.3%; Score 77; DB 1; Length 286;
Best Local Similarity 26.3%; Pred. No. 5.57e-01;
Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 APASAPYSSDTPCCFAYIAR-PLP 43
QY 20 SSAPRKLNTVNGCCFATVTRPLS 45

RESULT 12
ID RIP2_MOMBA STANDARD; PRT; 286 AA.
AC F29339;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
DE RIBOSOME-INACTIVATING PROTEIN MOMORDIN II PRECURSOR (RRNA
DE N-GLYCOSIDASE) (EC 3.2.2.22).
OS MOMORDICA BALSAMINA (BITTER GOURD) (BALSAM PEAR).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
CN CUCURBITALES; CUCURBITACEAE; MOMORDICA.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SEED;
RX MEDLINE; 93027170.
RA "Momordin II, a ribosome inactivating protein from Momordica
RT balsamina, is homologous to other plant proteins.";
RL NUCLEIC ACIDS RES. 20:4662-4662(1992).
CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT ONE
CC SPECIFIC ADENOSINE ON THE 28S RRNA.
CC -1- SIMILARITY: TO OTHER BACTERIAL AND PLANTS RIBOSOME-INACTIVATING
CC PROTEINS. BELONGS TO TYPE 1 RIP.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
-----
DR EMBL; X66035; G387;
DR PIR; S26596; S26596.
DR PROSITE; PS00215; MITOCH_CARRIER; 3.
DR PFAM; PF00153; mito_carr; 1.
RW MITOCHONDRIUM; TRANSMEMBRANE; TRANSPORT; REPEAT.
FT REPEAT 122 217
FT REPEAT 121 121
FT REPEAT 218 330
SQ SEQUENCE 330 AA; 36085 MW; 428468D4 CRC32;

Query Match 17.3%; Score 77; DB 1; Length 330;
Best Local Similarity 48.1%; Pred. No. 5.57e-01;
Matches 13; Conservative 3; Mismatches 8; Indels 3; Gaps 3;

Db 36 RSFLAGGIAGCC-AKTTPVAPLDVRKVL 61
QY 25 RKLTN-VTGCCFATVTRPLSN-KVL 49

RESULT 14
ID N55M_RHUN STANDARD; PRT; 606 AA.
AC Q96069;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 5 (EC 1.6.5.3).
GN MTND5 OR NDS OR NADH5.
OS RHINOCEROS UNICORNIS (GREATER INDIAN RHINOCEROS).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
```



```
Search completed: Wed, Dec 22 10:20:19 1999
Job time : 0 secs.
```

 W P S R L H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:20:37 1999; MasPar time 10.31 Seconds
 312.331 Million cell updates/sec
 Tabular output not generated..

Title: >US-09-209-961-7
 Description: (1-59) from US09209961.pep
 Perfect Score: 445
 Sequence: 1 MKCTLVFQSRFCIFLTKFS.....TRIPLSNKVLTAVDRLRCP 59

Scoring table: PAM 150
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrmb19
 1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 35.655; Variance 56.771; scale 0.628

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	445	100.0	59	14	092287 P6.6.	1.02e-84
2	418	93.9	59	14	056125 STRAIN PMWS PCV, COMPL	6.24e-78
3	89	20.0	1206	11	035404 SYNAPTOJANIN 2 (FRAGME	2.13e-02
4	89	20.0	1248	11	055207 SYNAPTOJANIN II	2.13e-02
5	84	18.9	402	14	095484 ENVELOPE POLYPROTEIN (1.48e-01
6	84	18.9	454	5	045869 T27F6.5 PROTEIN.	1.48e-01
7	84	18.9	601	14	064984 ENV-PR95 POLYPROTEIN P	1.48e-01
8	84	18.9	650	5	044236 ENDOSTYLE-SPECIFIC.	1.48e-01
9	83	18.7	141	14	084663 GENOME, PARTIAL SEQUEN	2.16e-01
10	82	18.4	319	2	03037 SHIGA-LIKE TOXIN II SU	3.15e-01
11	80	18.0	323	5	029269 F28A12.2 PROTEIN.	6.62e-01
12	80	18.0	345	14	035473 (RAV-0 SUBGROUP E) GP-	6.62e-01
13	79	17.8	286	10	041257 MAP30 (FRAGMENT)	9.56e-01
14	79	17.8	311	6	077792 UNCOUPLING PROTEIN 3.	9.56e-01
15	79	17.8	326	10	065355 GAMMA-GLUTAMYL HYDROLA	9.56e-01
16	79	17.8	592	2	051094 PEPTIDASE, PUTATIVE.	9.56e-01
17	78	17.5	295	5	039744 SIMILAR TO MOUSE CREB-	1.38e+00
18	78	17.5	313	14	036215 POLYPROTEIN (FRAGMENT)	1.38e+00
19	78	17.5	319	2	047643 (EC 3.2.2.22) (RRNA N-	1.38e+00
20	77	17.3	91	4	043546 RANTES PRECURSOR.	1.98e+00

21	77	17.3	125	6	077766 M3 MUSCARINIC ACETYLCH	1.98e+00
22	77	17.3	260	14	065555 UL34.	1.98e+00
23	77	17.3	501	5	016923 F25G6.7 PROTEIN.	1.98e+00
24	77	17.3	4976	2	087314 FXBC.	1.98e+00
25	76	17.1	304	5	045882 W03B1.7 PROTEIN.	2.83e+00
26	76	17.1	320	2	046050 (EC 3.2.2.22) (RRNA N-	2.83e+00
27	76	17.1	440	2	025161 NIFS-LIKE PROTEIN.	2.83e+00
28	76	17.1	443	5	091351 CODED FOR BY C. ELEGAN	2.83e+00
29	76	17.1	648	10	041717 ACETOYLACTATE SYNTHASE	2.83e+00
30	76	17.1	684	4	016707 PROSTATE-SPECIFIC TRAN	2.83e+00
31	76	17.1	845	5	091184 SIMILAR TO C. ELEGANS	2.83e+00
32	75	16.9	271	14	085474 ENVELOPE PROTEIN (FRAG	4.03e+00
33	75	16.9	301	5	022299 T07C5.4 PROTEIN.	4.03e+00
34	75	16.9	319	2	047642 (EC 3.2.2.22) (RRNA N-	4.03e+00
35	75	16.9	319	2	047636 (EC 3.2.2.22) (RRNA N-	4.03e+00
36	75	16.9	335	14	085475 ENVELOPE PROTEIN (FRAG	4.03e+00
37	75	16.9	454	5	090792 D2030.7 PROTEIN.	4.03e+00
38	75	16.9	577	14	007453 ENV POLYPROTEIN (CONTA	4.03e+00
39	75	16.9	595	14	003819 ENV POLYPROTEIN (CONTA	4.03e+00
40	75	16.9	607	14	083134 PROTEIN KINASE C, MU (4.03e+00
41	75	16.9	918	11	062101 ANKYRIN HOMOLOG.	4.03e+00
42	75	16.9	936	10	023292 1-EVIDENCE-PREDICTED B	4.03e+00
43	75	16.9	990	5	046086 SYNAPTOJANIN 2 ISOFORM	4.03e+00
44	75	16.9	1145	11	088399 POLYPROTEIN.	4.03e+00
45	75	16.9	2842	14	036178	4.03e+00

ALIGNMENTS

RESULT	ID	PRELIMINARY:	PRT:	59 AA.
AC	092287			
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	P6.6.			
OC	PORCINE CIRCOVIRUS.			
OS	VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.			
R4	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=412;			
RA	WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;			
RL	"Emergence of a new porcine circovirus.";			
DR	SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.			
DR	EMBL: AF085695; G3668369.			
SQ	SEQUENCE 59 AA; 6603 MW; 016C3C5F CRC32;			
Query Match 100.0%; Score 445; DB 14; Length 59;				
Best Local Similarity 100.0%; Pred. No. 1.02e-84;				
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
CU	1 MKCTLVFQSRFCIFLTKFSASPRKFLTNVTCGFATVTRIPLSNKVLTAVDRLRCP 59			
CU				
CU	1 MKCTLVFQSRFCIFLTKFSASPRKFLTNVTCGFATVTRIPLSNKVLTAVDRLRCP 59			
CU				
RESULT	ID	PRELIMINARY:	PRT:	59 AA.
AC	056125			
DT	01-JUN-1998 (TREMBLREL. 06, CREATED)			
DT	01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)			
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)			
DE	STRAIN PMWS PCV, COMPLETE GENOME.			
OC	PORCINE CIRCOVIRUS.			
OS	VIRUSES; SSDNA VIFUSES; CIRCOVIRIDAE; CIRCOVIRUS.			
R4	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=PMWS PCV;			
RA	MEDLINE, 9824172.			
RA	HAMEL A.L., LIN L.L., NAYAK G.P.S.;			
RT	"Nucleotide sequence of porcine circovirus associated with			
RT	postweanair, multisystemic wasting syndrome in pigs.";			
PL	J. VIROL. 72:5262-5267(1998).			

```

[2]
RP SEQUENCE FROM N.A.
RC STRAIN-PWS PCV;
RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RX MEDLINE; 98418498.
RA MEEHAN B.M., MCNEILLY F.M., TODD D.G., KENNEDY S., JEWELL V.,
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
RT "Characterisation of novel circovirus DNAs associated with wasting
RT syndromes in pigs.";
RL J. GEN. VIROL. 79:2171-2199(1998).
[4]
RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RA MEEHAN B.M.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF027217; G3598835;
DR EMBL; AF055394; G3598835;
DR EMBL; AF055391; G3598800;
DR EMBL; AF055392; G3598812;
DR EMBL; AF055393; G3598824;
SQ SEQUENCE 59 AA; 6516 MW; 11076BA8 CRC32;

Query Match 93.9%; Score 418; DB 14; Length 59;
Best Local Similarity 94.9%; Pred. No. 6.24e-78;
Matches 56; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Db 1 MTCTLVFQSFRCIFPLTFKSSASPRKFLTNVTCGCSATVTRPLSNKVLAVDRSLRCP 59
QY 1 MKCTLVFQSFRCIFPLTFKSSASPRKFLTNVTCGCFATVTRIPLSNKLAVDRSLRCP 59

RESULT 3
ID Q35404 PRELIMINARY; PRT; 1206 AA.
AC Q35404;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JUN-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SYNAPTOJANIN 2 (FRAGMENT).
OS MUS MUSCULUS (MOUSE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; MUS.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RX MEDLINE; 98113197.
RA KHVOICHEV M., SUDHOF T.C.;
RT "Developmentally regulated alternative splicing in a novel
RT synaptojanin.";
RL J. BIOL. CHEM. 273:2306-2311(1998).
DR EMBL; AF026123; G3478621;
FT NON_TER 1
SQ SEQUENCE 1206 AA; 133879 MW; C63591C9 CRC32;

Query Match 20.0%; Score 89; DB 11; Length 1206;
Best Local Similarity 46.9%; Pred. No. 2.13e-02;
Matches 15; Conservative 7; Mismatches 8; Indels 2; Gaps 1;

Db 21 CLGELSLKSGVPLSLFLVLTGCG--TSVGRIP 50
QY 12 CIFPLTFKSSASPRKFLTNVTCGCFATVTRIP 43

RESULT 4
ID Q55207 PRELIMINARY; PRT; 1248 AA.
AC Q55207;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE SYNAPTOJANIN II.

```

```

OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
OC SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 98049546.
RA NEMOTO Y., ARRIBAS M., HAFNER C., DECAMILLI P.;
RT "Synaptojanin 2, a novel synaptojanin isoform with a distinct
RT targeting domain and expression pattern.";
RL J. BIOL. CHEM. 272:30817-30821(1997).
DR EMBL; U90312; G2708493;
SQ SEQUENCE 1248 AA; 138274 MW; 4E57DB9E CRC32;

Query Match 20.0%; Score 89; DB 11; Length 1248;
Best Local Similarity 46.9%; Pred. No. 2.13e-02;
Matches 15; Conservative 7; Mismatches 8; Indels 2; Gaps 1;

Db 63 CLGELSLKSGVPLSLFLVLTGCG--TSVGRIP 92
QY 12 CIFPLTFKSSASPRKFLTNVTCGCFATVTRIP 43

RESULT 5
ID Q85484 PRELIMINARY; PRT; 402 AA.
AC Q85484;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ENVELOPE POLYPROTEIN (FRAGMENT).
GN ENV.
OS ROUS SARCOMA VIRUS.
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
[1]
RP SEQUENCE FROM N.A.
RC STRAIN-SCHMIDT-RUBBIN SUBGROUP D;
RX MEDLINE; 88063012.
RA BOVA C.A., OLSEN J.C., SWANSTROM R.;
RT "The avian retrovirus env gene family: molecular analysis of host
RT range and antigenic variants.";
RL J. VIROL. 62:75-83(1988).
DR EMBL; M22730; G825438;
KW POLYPROTEIN; ENVELOPE PROTEIN.
FT NON_TER 1
FT CHAIN 53 >402 GLYCOPROTEIN 85.
FT NON_TER 402 402
SQ SEQUENCE 402 AA; 43565 MW; 9279511E CRC32;

Query Match 18.9%; Score 84; DB 14; Length 402;
Best Local Similarity 35.3%; Pred. No. 1.48e-01;
Matches 12; Conservative 10; Mismatches 11; Indels 1; Gaps 1;

Db 301 PISFTVNTGSGN-LGNVSGCCGEATILPLGAW 333
QY 15 PLTFKSSASPRKFLTNVTCGCFATVTRIPLSNKV 48

RESULT 6
ID Q45869 PRELIMINARY; PRT; 454 AA.
AC Q45869;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE T27F6.5 PROTEIN.
GN T27F6.5.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
[1]
RP SEQUENCE FROM N.A.
RA DOBSON R.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.

```

RA MEDLINE; 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
 RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans*;
 RL NATURE 368:32-38(1994).
 DR EMBL; 282060; E1349861;
 SQ SEQUENCE 454 AA; 50593 MW; 94C7F3FD CRC32;

Query Match 18.9%; Score 84; DB 5; Length 454;
 Best Local Similarity 35.4%; Pred. No. 1.48e-01;
 Matches 17; Conservative 10; Mismatches 17; Indels 4; Gaps 4;

Db 268 FHGKYSALCAKFGKQPLDMCCFGVIGVTRLLPAIDLVSVDKALR 315
 QY 14 FPLTFKSSAS-PRKFLTN-VTCCFAT-VTRI-PLSNKVLTVADRSLR 57

RESULT 7
 ID 064984; PRELIMINARY; PRT; 601 AA.
 AC 064984;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ENV-PR95 POLYPROTEIN PRECURSOR.
 GN ENV.
 OS ROUS SARCOMA VIRUS
 OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PRAGUE;
 RX MEDLINE; 83155662.
 RA SCHWARTZ D.E., TIZARD R., GILBERT W.,
 RT "Nucleotide sequence of Rous sarcoma virus*";
 RL CELL 32:853-869(1983).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PRAGUE;
 RX MEDLINE; 85124605.
 RA BROOME S., GILBERT W.,
 RT "Rous sarcoma virus encodes a transcriptional activator*";
 RL CELL 40:537-546(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA PETROPOULOS C.J.;
 RL (IN) COFFIN J.M. (EDS.); RETROVIRUSES:757-757;
 RL COLD SPRING HARBOR LABORATORY PRESS, NY, USA (1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA CHAPPEY C.;
 RL SUBMITTED (NOV-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; J02342; G210174;
 DR EMBL; AF033808; G2801462;
 KW POLYPROTEIN.
 FT CHAIN 1 175 POTENTIAL.
 FT CHAIN 63 601 POTENTIAL.
 FT CHAIN 404 601 POTENTIAL.
 SQ SEQUENCE 601 AA; 65316 MW; 6A2AF544 CRC32;

Query Match 18.9%; Score 84; DB 14; Length 601;
 Best Local Similarity 35.5%; Pred. No. 1.48e-01;
 Matches 11; Conservative 10; Mismatches 9; Indels 1; Gaps 1;

Db 302 PFSFTANCTGSN-LGNVSGCCGEPITILPLG 331
 QY 14 FPLTFKSSAS-PRKFLTN-VTCCFAT-VTRI-PLSNKVLTVADRSLR 57

QY 15 PLTFKSSASPRKFLTNVTGCCFATVTRIPLS 45
 RESULT 8
 ID 044236; PRELIMINARY; PRT; 650 AA.
 AC 044236;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE ENDSTYLE-SPECIFIC.
 OS CIONA INTESTINALIS.
 CC EUKARYOTA; METAZOA; CHORDATA; UROCHORDATA; ASCIDIACEA; PHLEBOBRANCHIA;
 CC CIONIDAE; CIONA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-ENDSTYLE;
 RA OGASAWARA M., SATOH N.,
 RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AB010895; D1025751;
 SQ SEQUENCE 650 AA; 75483 MW; 077A54DA CRC32;

Query Match 18.9%; Score 84; DB 5; Length 650;
 Best Local Similarity 31.6%; Pred. No. 1.48e-01;
 Matches 12; Conservative 10; Mismatches 12; Indels 4; Gaps 4;

Db 344 YKTRYCYVRYCLOPPKFCFNPKY-TGLKTCVFPVAVR 380
 QY 7 FQSREC-I-FPL-TFKSSASPRKFLTNVTGCCFATVTR 41

RESULT 9
 ID 084563; PRELIMINARY; PRT; 141 AA.
 AC 084563;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE GENOME, PARTIAL SEQUENCE.
 GN A349L.
 OS PARAMECIUM BURSARIA CHLORELLA VIRUS 1 (PBCV-1).
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; PHYCODNAVIRIDAE; PHYCODNAVIRUS.
 RN [1]
 RP SEQUENCE FROM W.A.
 RX MEDLINE; 95133167.
 RA LU Z., LI Y., ZHANG Y., KUTISH G.F., ROCK D.L., VAN ETEN J.L.;
 RT "Analysis of 45 kb of DNA located at the left end of the chlorella
 RT virus PBCV-1 genome";
 RL VIROLOGY 206:339-352(1995).
 DR EMBL; U42580; G1181512;
 SQ SEQUENCE 141 AA; 16438 MW; DCEBBA2E CRC32;

Query Match 18.7%; Score 83; DB 14; Length 141;
 Best Local Similarity 38.1%; Pred. No. 2.16e-01;
 Matches 8; Conservative 9; Mismatches 3; Indels 1; Gaps 1;

Db 5 FDARI/CLYSLYC/RTSGERRR 25
 QY 7 FQSRJIFPLT-FKSSASPRK 26

RESULT 10
 ID 003037; PRELIMINARY; PRT; 319 AA.
 AC 003037;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE SHIGA-LIKE TOXIN II SUBUNIT A PRECURSOR (VEROTOXIN 2 SUBUNIT A)
 DE (RNA N-GLYCOSIDASE) (EC 3.2.2.22).
 GN SLT-113.
 OS ESCHERICHIA COLI.
 CC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 CC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN-0X3:H21;
RA PATON A.W., PATON J.C., HEUZENROEDER M.W., GOLDWATER P.N.,
RL MANNING P.A.;
RM MICROB. PATHOG. 13:225-236(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-0111:H-;
RX MEDLINE; 94018566.
RA PATON A.W., PATON J.C., MANNING P.A.;
RT "Polymerase chain reaction amplification, cloning and sequencing of
RL variant Escherichia coli Shiga-like toxin type II operons";
RM MICROB. PATHOG. 15:77-82(1993).
CC -!- FUNCTION: THE SUBUNIT A IS RESPONSIBLE FOR INHIBITING PROTEIN
CC SYNTHESIS THROUGH THE CATALYTIC INACTIVATION OF 60S RIBOSOMAL
CC SUBUNITS.
CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF THE N-GLYCOSIDIC BOND AT
CC ONE SPECIFIC ADENOSINE ON THE 28S RRNA.
CC -!- SUBUNIT: THE SHIGA-LIKE TOXINS CONTAIN A SINGLE A SUBUNIT AND
CC MULTIPLE COPIES OF A B SUBUNIT.
CC -!- SIMILARITY: TO ALPHA CHAINS OF SHIGA TOXIN, SHIGA-LIKE TOXIN I,
CC TO THE A SUBUNIT OF THE PLANT TOXIN RICIN, AND TO OTHER PLANT
CC RIBOSOME-INACTIVATING PROTEINS.
DR EMBL; X65949; G49090;
DE EMBL; L11078; G304951;
DR PFAM; PF00161; RIP; 1;
KW HYDROLASE; GLYCOSIDASE; TOXIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 319 SHIGA-LIKE TOXIN II SUBUNIT A.
FT TRANSMEM 232 267 POTENTIAL.
FT SITE 188 188 ESSENTIAL FOR ENZYMIC ACTIVITY.
FT SITE 192 192 ESSENTIAL FOR ENZYMIC ACTIVITY.
FT ACT_SITE 189 189 BY SIMILARITY.
FT DISULFID 263 282 POTENTIAL.
FT VARIANT 198 198 R -> G (IN STRAIN 0111:H-).
SQ SEQUENCE 319 AA; 35709 MW; F0331EA CRC32;
Query Match 18.4%; Score 82; DB 2; Length 319;
Best Local Similarity 36.7%; Pred. No. 3.15e-01;
Matches 11; Conservative 7; Mismatches 10; Indels 2; Gaps 2;
Db 1 MKCIL-FKWILCLL-LGFSVSVSREFMID 28
QY 1 MKCTLVFQSRFCIFPLTFKSSASPRKFLTN 30
RESULT 11
ID Q22969 PRELIMINARY; PRT; 323 AA.
AC Q22969;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-JUN-1998 (TREMREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE F28A12.2 PROTEIN.
GN F28A12.2
OS CAENORHABDITIS ELGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAYELLO A., FULTON L.,
RA GARDNER A., GREEN P., HANKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., McMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL NATURE 368:32-38(1994).

AN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA SAMMONS L., MURRAY J.;
RL SUBMITTED (JUL-1996) TO EMBL/GENBANK/DBDJ DATA BANKS.
DR EMBL; U64851; G2911709;
SQ SEQUENCE 323 AA; 35770 MW; FCA220D4 CRC32;
Query Match 18.0%; Score 80; DB 5; Length 323;
Best Local Similarity 34.3%; Pred. No. 6.62e-01;
Matches 12; Conservative 8; Mismatches 13; Indels 2; Gaps 2;
Db 10 IFNRFNFEIPFHSGIAPGRRTMNSIFCCFNS 44
QY 6 VFQSRFCIF-PLTFKSSASP-RKFLTNVTCGFAT 38
RESULT 12
ID Q85473 PRELIMINARY; PRT; 345 AA.
AC Q85473;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE (RAY-0 SUBGROUP E) GP-85 (ENV) (FRAGMENT).
OS ROUS SARCOMA VIRUS.
OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; AVIAN TYPE C RETROVIRUSES.
RN [1]
RP SEQUENCE OF 270-345 FROM N.A.
RA DORNER A.J.;
RL SUBMITTED (JUL-1988) TO EMBL/GENBANK/DBDJ DATA BANKS.
RN [2]
RP SEQUENCE OF 1-269 FROM N.A.
RX MEDLINE; 85083270.
RA DORNER A.J., STOEY J.P., COFFIN J.M.;
RT "Molecular basis of host range variation in avian retroviruses";
RL J. VIROL. 53:32-39(1985).
DR EMBL; M12172; G559652;
FT NON_TER 1 1
SQ SEQUENCE 345 AA; 37411 MW; B26AE001 CRC32;
Query Match 18.0%; Score 80; DB 14; Length 345;
Best Local Similarity 37.9%; Pred. No. 6.62e-01;
Matches 11; Conservative 10; Mismatches 7; Indels 1; Gaps 1;
Db 244 PFSETANCTASN-LGNVSGCGCKTITILP 271
QY 15-PLTFKSSASPRKFLTNVTCGFATVTRIP 43
RESULT 13
ID Q41257 PRELIMINARY; PRT; 286 AA.
AC Q41257;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMREL. 09, LAST ANNOTATION UPDATE)
DE MAP30 (FRAGMENT).
GN MAP30.
OS MOMORDICA CHARANTIA (BITTER GOURD) (BALSAM PEAR).
OC EUKARYOTA; VIRIDIPLANTAE; STREPTOPHYTA; TRACHEOPHYTA;
OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS; ROSIDAE;
OC CUCURBITALES; CUCURBITACEAE; MOMORDICA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95394347.
RA LEE-HUANG S., HUANG P.L., CHEN H.C., HUANG P.L., BOURINBAIR A.,
RA HUANG H.I., KUNG H.F.;
RT "Anti-HIV and anti-tumor activities of recombinant MAP30 from bitter
RL melon";
RL GENE 161:151-156(1995).
DR EMBL; S79450; G110597;
DR PFAM; PF00161; RIP; 1.
DR MENDEL; 15516; MOMCH; Rip; mn15516.
FT NON_TER 286

W P S R E H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:22:13 1999; MasPar time 13.84 Seconds
Tabular output not generated. 159.807 Million cell updates/sec

Title: >US-09-209-961-9
Description: (1-104) from US09209961.pep
Perfect Score: 807
Sequence: 1 MVTIPPLVFRWFPVCGFRVC.....HSSRQVTPLSLRSRSTFNK 104
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266608 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 29.035; Variance 113.207; scale 0.256

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query		Match	Length	DB ID	Description	Pred. No.
	Score	Match					
1	89	11.0	397	28	W34307	Human neurotactin.	9.65e+00
2	89	11.0	397	26	W23345	Novel human CX3C 397	9.65e+00
3	89	11.0	787	32	W37912	Homo sapiens Smoothen	9.65e+00
4	89	11.0	793	32	W37911	Rattus norvegicus Smo	9.65e+00
5	86	10.7	534	26	W36137	A. thaliana inorganic	1.69e+01
6	86	10.7	534	27	W32302	Arabidopsis thaliana	1.69e+01
7	84	10.4	440	2	P70134	Natural recombinant h	2.45e+01
8	84	10.4	440	25	W24789	Human lecithin-choles	2.45e+01
9	81	10.0	111	26	W23344	Novel human CX3C 111	4.23e+01
10	81	10.0	159	12	R67752	Tomv coat protein.	4.23e+01
11	81	10.0	372	28	W44284	N-methylamino acid ox	4.23e+01
12	81	10.0	956	26	W27086	HCW UL105 open read	4.23e+01
13	81	10.0	956	26	W27083	Human cytomegalovirus	4.23e+01
14	80	9.9	348	12	R63788	POU domain transcript	5.07e+01
15	79	9.8	227	39	W95515	Fatty acid desaturase	6.08e+01
16	79	9.8	227	38	W85125	A putative delta-5 de	6.08e+01

SUMMARIES

17 79 9.8 227 37 W84146
18 75 9.8 345 31 W56756
19 79 9.8 404 7 R34581
20 79 9.8 435 12 R58689
21 79 9.8 445 10 R54782
22 79 9.8 448 11 R57200
23 79 9.8 450 19 R99427
24 79 9.8 451 29 W38426
25 79 9.8 439 34 W62858
26 78 9.7 439 36 W74484
27 77 9.5 109 34 W29636
28 77 9.5 159 12 R60611
29 77 9.5 456 28 W41695
30 77 9.5 482 10 R53461
31 77 9.5 482 8 R41659
32 77 9.5 482 27 W49813
33 77 9.5 824 34 W68093
34 76 9.4 164 21 W11787
35 76 9.4 207 22 W15288
36 75 9.3 189 13 R66605
37 75 9.3 189 13 W14913
38 75 9.3 221 34 W38578
39 75 9.3 319 18 W00628
40 75 9.3 450 19 R99426
41 75 9.3 534 30 W45748
42 75 9.3 631 18 R77896
43 75 9.3 631 30 W54128
44 75 9.3 631 28 W08970
45 75 9.3 679 1 P95056

Dictyostellium discoide
Seroconin SH77 recept
Mammalian serotonin r
Rat REC20 serotonin r
Human brain serotonin
Rat 5HT6 receptor.
Human lymphocyte spec
Human multiple myelom
Peniophora lycii 6-ph
Peniophora phytase po
Human secreted protei
Tobamovirus coat prot
Rabbit LPS-binding pr
Lapine polysaccharide
Lapine gram-negative
Rabbit lipopolysaccha
Human neuronal PAS do
Early onset Alzheimer
Mouse oligodendrocyte
Flower style-specific
Tomato S-ribonuclease
S. pneumoniae stage V
Protein kinase #1 con
Murine lymphocyte spe
Arabidopsis thaliana
Bacterial transferrin
H. influenzae strain
Amino acid sequence o
Swine parvovirus B ge

ALIGNMENTS

1
W34307 standard; Protein: 397 AA.
W34307:
11-MAY-1998 (first entry)
Human neurotactin.
KW Neurotactin; human; chemokine; inflammatory disease; inflammation;
KW neutrophil; chemotaxis; myelogenous leukaemia; polycythemia vera;
KW hypermegakaryocytopenic disorder; chemoprotective;
KW radioprotective; therapy; diagnosis; Bardet-Biedl syndrome.
OS Homo sapiens.
TS Key Location/Qualifiers
FT Peptide 1..221
FT Domain 22..341
FT Domain /note= "extracellular domain"
FT Domain 22..92
FT Domain /note= "chemokine-like domain (Claim 5)"
FT Domain 342..362
FT Domain /note= "transmembrane domain"
FT Domain 363..397
FT Domain /note= "cytoplasmic domain"
W09742224-11-
13-NOV-1997:
06-MAY-1997: U07737.
05-MAY-1997: US-851160.
07-MAY-1996: US-643798.
(MILL-) MILLENNIUM PHARM INC.
PAN Y:
WPI: 97-556911/51.
N-PSDB: T93210.
DNA encoding murine and human neurotactin - useful for diagnosis and treatment of inflammatory disorders
Claim 5; Page 76-78; 101pp; English.
This protein is human neurotactin, a novel chemokine that stimulates chemotaxis of neutrophils. Its amino acid sequence was deduced from the isolated neurotactin gene (see T93210). Neurotactin has a N-terminal chemokine-like domain, but its overall structure distinguishes it from other chemokines. It is highly expressed in the normal human brain, and may play a role in Bardet-Biedl syndrome. Neurotactin polypeptides, including soluble neurotactin

21-000-1597.
23-JAN-1997; U00293.
16-MAY-1996; US-649006.
24-JAN-1996; US-590828.
(SCHE) SCHERING CORP.
Bazan JF, Schall TJ, Zlotnik A;
WPI; 97-393687/36.
N-PSDB; T64943.
New isolated mammalian CX3C chemokine genes - used to develop
products for the diagnosis and treatment of disorders involving CX3C
chemokine misregulation
Claim 18; Pages 75-76; 88pp; English.
PS

RESULT	3	
ID	W37912 standard; Protein; 787 AA.	
AC	W37912;	
DE	11-SEP-1998 (first entry)	
DE	Homo sapiens Smoothed protein.	
KW	Smoothened; treatment: cancer; vertebrae; vSmo; tissue;	
KW	regeneration; skin; lung; muscle; neural; bone; gut; disease;	
KW	ageing; trauma; detection; diagnosis; drug screening.	
OS	Homo sapiens.	
PN	WO9814475-A1.	
PD	09-APR-1998.	
PD	29-SEP-1997; U17433.	
PR	30-SEP-1996; US-720484.	
PA	(GETH) GENENTECH INC.	
PI	De SAUVAGE FJ, Rosenthal A, Stone DM;	
DR	WPI; 98-240021/21.	
DR	N-PSDB; V29125.	
PT	New isolated vertebrae Smoothened proteins - is used to develop	
PT	products for treatment of cancers or for use in inducing	
PT	regeneration of tissue damaged by disease, ageing or trauma	
PS	Claim 5; Fig 4; 58pp; English.	
CC	The sequence is that of human Smoothened protein (Smo).	
CC	Smo interacts with Hedgehog and Patched signalling molecules	
CC	involved in cell proliferation and differentiation. The products	
CC	can be used to develop agents to provide protection from conditions	
CC	associated with constitutive activity of Smo or Hedgehog, including	
CC	some forms of cancer that may result from e.g. basal cell carcinoma,	
CC	basal cell nevus syndrome and pancreatic carcinoma. Smo agonists	
CC	can be used to induce the formation of, or enhance or stimulate	
CC	tissue regeneration, such as regeneration of skin tissue, lung	
CC	tissue, muscle (such as heart or skeletal muscle), neural tissue	
CC	(such as serotonergic neurons, motoneurons or striatal neurons),	
CC	bone tissue or gut tissue, which may be damaged by disease,	
CC	ageing or trauma. The products can also be used for detection,	
CC	diagnosis and drug screening.	
SQ	Sequence 787 AA;	
	SQ	

Query Match 11.0%; Score 89; DB 32; Length 787;
 Best Local Similarity 27.9%; Pred. No. 9.65e+00;
 Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 457 flafgvlitfchdydfnqae-wersfrdyvlcqanvtgl 498
 QY 39 YIGFPITLLHFAHFKQSPAEIFDKRYV-LLCNGHONPAL 80

RESULT 4
 ID W37911 standard; Protein; 793 AA.

AC W37911.1998 (first entry)
 DE Rattus norvegicus Smoothed protein.
 KW Smoothed; treatment; cancer; vertebrae; vsmo; tissue;
 KW regeneration; skin; lung; muscle; neural; bone; gut; disease;
 KW ageing; trauma; detection; diagnosis; drug screening.
 OS Rattus norvegicus.
 PN W09814475-A1.
 PD 09-APR-1998.
 PF 29-SEP-1997; U17433.
 PR 30-SEP-1996; US-720484.
 PA (GETH) GENENTECH INC.
 PI De SAVAGE FJ, Rosenthal A, Stone DM;
 DR WPI: 98-240021/21.
 DR N-PSDB; V29124.
 PT New isolated vertebrate Smoothed proteins - is used to develop
 PT products for treatment of cancers or for use in inducing
 PT regeneration of tissue damaged by disease, ageing or trauma
 PS Claim 6; Fig 1; 58pp; English.
 CC The sequence is that of rat Smoothed protein (Smo).
 CC Smo interacts with Hedgehog and Patched signalling molecules.
 CC involved in cell proliferation and differentiation. The products
 CC can be used to develop agents to provide protection from conditions
 CC associated with constitutive activity of Smo or Hedgehog, including
 CC some forms of cancer that may result from e.g. basal cell carcinoma,
 CC basal cell nevus syndrome and pancreatic carcinoma. Smo agonists
 CC can be used to induce the formation of, or enhance or stimulate
 CC tissue regeneration, such as regeneration of skin tissue, lung
 CC tissue, muscle (such as heart or skeletal muscle), neural tissue
 CC (such as serotonergic neurons, motoneurons or striatal neurons),
 CC bone tissue or gut tissue, which may be damaged by disease,
 CC ageing or trauma. The products can also be used for detection,
 CC diagnosis and drug screening.
 SQ Sequence 793 AA;

Query Match 11.0%; Score 89; DB 32; Length 793;
 Best Local Similarity 27.9%; Pred. No. 9.65e+00;
 Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 461 flafgvlitfchdydfnqae-wersfrdyvlcqanvtgl 502
 QY 39 YIGFPITLLHFAHFKQSPAEIFDKRYV-LLCNGHONPAL 80

RESULT 5
 ID W36137 standard; Protein; 534 AA.

AC W36137.1998 (first entry)
 DE A. thaliana inorganic phosphate transporter protein IPT4.
 DE Promoter; inorganic phosphate transporter gene; IPT4; expression;
 KW regulation; transgenic plant; expression vector.
 OS Arabidopsis thaliana.
 PN J09252782-A.
 PD 30-SEP-1997.
 PF 25-MAR-1996; 094856.
 PR 25-MAR-1996; JP-094856.
 PA (MITS-) MITSUI BIOSCI INC. SHOKUBUTSU BIO KENKYUSHO.
 DR WPI: 97-530154/49.
 DR N-PSDB; V01455.
 PT A promoter for a phosphate transporter gene - useful for regulating
 PT heterologous gene expression in transgenic plants
 PS Example 2; Page 6-9; 12pp; Japanese.

CC This is the amino acid sequence of the Arabidopsis thaliana inorganic
 CC phosphate transporter gene IPT4. The promoter region (bases 1-1875; see
 CC also V01454) of the encoding gene can be used to regulate the level of
 CC expression of a heterologous gene in a plant by introducing, into a
 CC plant, a vector having the heterologous gene linked downstream region
 CC from the promoter and then regulating extracellular phosphate levels.
 SQ Sequence 534 AA;

Query Match 10.7%; Score 86; DB 26; Length 534;

Best Local Similarity 42.1%; Pred. No. 1.69e+01;
 Matches 24; Conservative 9; Mismatches 15; Indels 9; Gaps 5;

Db 385 flalaipynhwhkenrigfvmysltffanfgnattfvvpaefparfirst-chg 443
 QY 26 FAFPTTP--RWPNEVYIGPPI--TLHFFPAHF--OK---FSQPAEIFDKRYVLLCNG 74

RESULT 6
 ID W32302 standard; Protein; 534 AA.

AC W32302.1998 (first entry)
 DE Arabidopsis thaliana inorganic phosphate transporter 4.
 KW Columbia strain; inorganic phosphate transporter 4; IPT4;
 KW accelerated phosphate uptake; tobacco plant.
 OS Arabidopsis thaliana.
 PN W09735984-A1.
 PD 02-OCT-1997.
 PF 24-MAR-1997; J00975.
 PR 25-MAR-1996; JP-094790.
 PA (MITS-) MITSUI PLANT BIOTECHNOLOGY RES INST.
 PI Mitsukawa N, Okumura S, Shibata D, Shirano Y;
 DR WPI: 97-489647/45.
 DR N-PSDB; T91322, T91325.
 PT DNA encoding a plant phosphate transporter protein - useful for
 PT producing e.g. tobacco plants with increased phosphate uptake and
 PT accelerated growth.
 PS Claim 2; Pages 72-74; 96pp; Japanese.
 CC The present sequence is Arabidopsis thaliana (Columbia strain)
 CC inorganic phosphate transporter 4 (IPT4), useful for accelerating
 CC the uptake of phosphate by plants, e.g. tobacco plants.
 SQ Sequence 534 AA;

Query Match 10.7%; Score 86; DB 27; Length 534;
 Best Local Similarity 42.1%; Pred. No. 1.69e+01;
 Matches 24; Conservative 9; Mismatches 15; Indels 9; Gaps 5;

Db 388 flalaipynhwhkenrigfvmysltffanfgnattfvvpaefparfirst-chg 443
 QY 26 FAFPTTP--RWPNEVYIGPPI--TLHFFPAHF--OK---FSQPAEIFDKRYVLLCNG 74

RESULT 7
 ID F70134 standard; Protein; 440 AA.

AC F70134.1991 (first entry)
 DE Natural recombinant human lecithin:cholesterol acyltransferase.
 DE Lecithin:cholesterol acyltransferase; congenital deficiencies;
 KW renal disease; hepatitis; heart disease polymorphisms.
 FH Key Location/Qualifiers
 FT peptide 25..440
 FT /label= mature lecithin:cholesterol acyltrans-
 FT ferase
 PN EP-222591-A.
 PD 20-MAY-1987.
 PF 05-NOV-1986; 308624.
 PR 08-NOV-1985; US-796473.
 PA (GETH) GENENTECH INC.
 PI Baer BW, Drayna DT, Lawn RM, Mclean JW;
 DR WPI: 87-137586/20.
 DR N-PSDB; N70191.
 PT New DNA encoding lecithin:cholesterol acyl-transferase - useful
 PT in prodn. of enzyme for therapeutic use and in hybridisation
 PT assays

PS Disclosure; fig 2; 20pp; English.

CC This natural human lecithin:cholesterol acyltransferase (LCAT) is administered therapeutically to patients having a deficiency of the enzyme, esp. in congenital deficiencies, end stage renal disease and hepatitis. Amino acid sequence variants can be produced, having improved properties eg enhanced oxidative stability. This DNA is useful in hybridisation assays for identifying heart disease linked restriction enzyme polymorphisms.

CC See also N70192-93.

SQ Sequence 440 AA;

Query Match 10.48; Score 84; DB 25; Length 440;
Best Local Similarity 29.38; Pred. No. 2.45e+01;
Matches 12; Conservative 13; Mismatches 11; Indels 5; Gaps 5;

Db 268 ri-tttspwfmfsmawpdedhvfistp-sf-nytgdfqrf 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
18 RVCKISPFPAFTPR-WPHEVIGFPTILLHPPAH-FQKF 56

RESULT 8

ID W24789 standard; Protein; 440 AA.

AC W24789;

DE 04-FEB-1998 (first entry)

DE Human lecithin-cholesterol acyltransferase (LCAT).

KW Human lecithin-cholesterol acyltransferase; LCAT; lecithin; cholesterol; Artherosclerosis; heart disease; stroke; heart attack; cholesterol reduction; peripheral vascular disease; Fish Eye Syndrome; myocardial infarction; Classic LCAT Deficiency Syndrome.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..24

FT /label= signal_peptide

FT Protein 25..440

FT /label= mature_peptide

PN W09717434-A2.

PD 15-MAY-1997.

PF 08-NOV-1996; U18159.

PR 09-NOV-1995; US-006400.

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Brewer HB, Hoeg JM, Santamarina-Fojo S;

DR WPI: 97-281024/25.

DR N-PSDB; T78853.

PT Treating atherosclerosis in a mammal, preferably a human - by increasing lecithin cholesterol acyl:transferase activity in mammal's serum to decrease cholesterol accumulation.

PS Claim 3; Fig 5; 48pp; English.

CC This sequence is of human lecithin-cholesterol acyltransferase (LCAT), which catalyses the transfer of fatty acid from the sn-2 position of lecithin to the free hydroxyl group of cholesterol. Although the sequence is known (McLean et al (1986), Nucleic Acids Research 14, pages 9397-9406), a new use has been discovered for this enzyme.

CC Artherosclerosis is a pathological condition of mammals characterised by the accumulation of cholesterol in the arteries, which leads to heart disease, strokes, heart attacks and peripheral vascular disease. The enzyme is used in a novel method of treating atherosclerosis, which involves increasing the level of LCAT activity, which then causes a decrease in the accumulation of cholesterol. The method and the products can be used for the prophylaxis and treatment of atherosclerosis, and associated heart disease, myocardial infarction, stroke and peripheral vascular disease, as well as individuals suffering from Fish Eye Syndrome (caused by LCAT deficiency) or Classic LCAT Deficiency Syndrome.

SQ Sequence 440 AA;

Query Match 10.48; Score 84; DB 25; Length 440;
Best Local Similarity 29.38; Pred. No. 2.45e+01;
Matches 12; Conservative 13; Mismatches 11; Indels 5; Gaps 5;

Db 268 ri-tttspwfmfsmawpdedhvfistp-sf-nytgdfqrf 305
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
18 RVCKISPFPAFTPR-WPHEVIGFPTILLHPPAH-FQKF 56

RESULT 9

ID W23344 standard; Protein; 111 AA.

AC W23344;

DE 30-MAR-1998 (first entry)

DE Novel human CX3C 111 amino acid chemokine.

KW Chemokine motif; CX3C chemokine; membrane bound CX3C chemokine; proadherent; T cell; monocyte; soluble CX3C chemokine; arthritis; proadhesive activity; inflammatory disease; compromised immune response; neurogenic inflammation.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..24

FT /label= signal_sequence

FT /note= "putative"

FT Protein 25..111

FT /label= mature_protein

FT Domain 25..100

FT /note= "chemokine domain"

PN W09727299-A1.

PD 31-JUL-1997.

PF 23-JAN-1997; U00293.

PR 16-MAY-1996; US-643006.

PR 24-JAN-1996; US-590828.

PA (SCHE) SCHERING CORP.

PI Bazan JF, Schall TJ, Zlotnik A;

DR WPI: 97-393687/36.

DR N-PSDB; T64942.

PT New isolated mammalian CX3C chemokine genes - used to develop products for the diagnosis and treatment of disorders involving CX3C chemokine misregulation

PS Claim 18; Page 73; 88pp; English.

CC The present sequence represents a protein that is part of a previously unknown class of chemokine motif containing molecules, which are designated the CX3C chemokines. The CX3C chemokines have three amino acids which separate the cysteines in the corresponding region of the chemokine motif. The membrane bound form of CX3C chemokine possesses proadherent properties for circulating T cells and monocytes. A secreted or soluble form, consisting of the chemokine domain and the stalk region, is able to inhibit this proadhesive activity. This suggests that the membrane bound form of CX3C chemokine may be a potent regulator of circulating leukocytes, and thus may be involved in various inflammatory diseases, e.g. arthritis. The soluble form may be used as a regulator of proadherence, especially in conditions of compromised immune response. The CX3C chemokines may also be involved in the pathogenesis of such CNS inflammatory disorders as multiple sclerosis, and other pathologies involving neurogenic inflammation. The products can be used to diagnose disorders associated with CX3C chemokine misregulation. They can also be used in the treatment of conditions associated with abnormal proliferation, e.g. cancerous conditions, or degenerative conditions. Abnormal proliferation, regeneration, degeneration, and atrophy may be modulated by appropriate therapeutic treatment using the products. The products can also be used in screening assays.

SQ Sequence 111 AA;

Query Match 10.0%; Score 81; DB 26; Length 111;
Best Local Similarity 23.7%; Pred. No. 4.23e+01;
Matches 14; Conservative 18; Mismatches 25; Indels 2; Gaps 2;

Db 43 ipvahlhyqngscskrailltetrlrfcdapkeqvkdamghldrqaaalcpkwr 101
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
42 FPITLL-HFPAHFQKFSQPAEFDKRVLLCNGHQPALQOGT-HSSRQVTPLSLR 98

RESULT 10

ID R67752 standard; Protein; 159 AA.

AC R67752;

DE 20-JUL-1995 (first entry)

DE ToMV coat protein.

KW Cucumber mosaic virus; tomato mosaic virus; minus-strand RNA; virus resistance; disease resistance; transgenic plant; cross protection; hypersensitive response; crop improvement;

KW tomato; Lycopersicon esculentum; tobacco; Nicotiana tabacum;

KW coat protein; CP; ToMV; CMV.

RESULT	l2	
ID	W27086	standard; Protein: 956 AA.
AC	W27086;	
DT	18-MAR-1998	(first entry)
DE	HCMV UL105 open reading frame translation product.	
DE	Human cytomegalovirus primase; HCMV UL105; open reading frame;orf:	
KW	screening; inhibitor; infection; diagnosis; ss.	
KW	OS	
OS	Homo sapiens.	
PN	GB2311069-A.	

```

: Query Match      10.0%; Score 81; DB 26; Length 956;
- Best Local Similarity 27.8%; Pred. No. 4.23e+01;
Matches 15; Conservative 15; Mismatches 22; Indels 2; Gaps

```

WQELH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd
MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:24:18 1999; MasPar time 4.00 Seconds
Tabular output not generated. 309.550 Million cell updates/sec

Title: >US-09-209-961-9
Description: (1-104) from US09209961.pep
Perfect Score: 807
Sequence: 1 MVTIPPLVFRWPPVCGFRVC.....HSSRQVTPLSLRSSSTFNK 104
Scoring table: PAM 150
Gap 11

Searched: 122461 seqs, 11912985 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: a-issued
1:SA_COMB 2:SB_COMB 3:PCT9_COMB 4:backfiles1
Statistics: Mean 27.055; Variance 110.616; scale 0.245

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match	Length DB ID	Description Pred. No.
1	88	10.9	775 1	US-07-603-Sequence 13, Applicati 6.28e+00
2	86	10.7	775 1	US-07-603-Sequence 14, Applicati 9.07e+00
3	81	10.0	159 2	US-08-553-Sequence 2, Applicatio 2.24e+01
4	80	9.9	348 1	US-08-035-Sequence 2, Applicatio 2.68e+01
5	80	9.9	348 1	US-08-504-Sequence 2, Applicatio 2.68e+01
6	79	9.8	345 2	US-08-745-Sequence 2, Applicatio 3.21e+01
7	79	9.8	365 3	PCT-US93-1Sequence 9, Applicatio 3.21e+01
8	79	9.8	365 2	US-08-428-Sequence 9, Applicatio 3.21e+01
9	79	9.8	404 2	US-08-428-Sequence 7, Applicatio 3.21e+01
10	79	9.8	404 3	PCT-US93-1Sequence 7, Applicatio 3.21e+01
11	79	9.8	422 1	US-07-996-Sequence 12, Applicati 3.21e+01
12	79	9.8	445 3	PCT-US93-1Sequence 2, Applicatio 3.21e+01
13	79	9.8	450 2	US-08-611-Sequence 25, Applicati 3.21e+01
14	78	9.7	3174 2	US-08-477-Sequence 3, Applicatio 3.83e+01
15	77	9.5	456 1	US-08-205-Sequence 4, Applicatio 4.57e+01
16	77	9.5	482 4	5245013-2Patent No. 5245013. 4.57e+01
17	77	9.5	775 1	US-07-603-Sequence 15, Applicati 4.57e+01
18	77	9.5	775 1	US-07-603-Sequence 16, Applicati 4.57e+01
19	77	9.5	824 2	US-08-816-Sequence 52, Applicati 4.57e+01
20	77	9.5	824 2	US-08-785-Sequence 7, Applicatio 4.57e+01
21	76	9.4	207 2	US-08-557-Sequence 2, Applicatio 5.45e+01
22	76	9.4	2291 2	US-08-286-Sequence 29, Applicati 5.45e+01
23	75	9.3	347 2	US-08-773-Sequence 3, Applicatio 6.49e+01


```
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC FRAGMENT TYPE: internal
SQ SEQUENCE 345 AA; 37857 MW; 648414 CN;

Query Match
Best Local Similarity 9.8%; Score 79; DB 2; Length 345;
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;

Db 213 ITLPL-FGWAQNVNDKVCILSQDFGTYIS-TAVAFYIPMSVLMFY 259
QY 2 VTIPPLVFRWFP-VGFRVCKISSPFAFTTPRPHNEVYIGFPITLLHF 49

RESULT 7
ID PCT-US93-10301-9 STANDARD; PRT; 365 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 9, Application PC/TUS9310301
Sequence 9, Application PC/TUS9310301
GENERAL INFORMATION:
APPLICANT: The United States of America, as represented by
APPLICANT: the Secretary of Health and Human Services
TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR
STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR
CITY: NEWPORT BEACH
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/428,243
FILING DATE: 18-SEP-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10301
FILING DATE: 27-OCT-1993
ATTORNEY/AGENT INFORMATION:
NAME: Altman, Daniel E.
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: NIH046.001VPC
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714-760-0404
TELEFAX: 714-760-9502
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 365 AA; 41165 MW; 680788 CN;

Query Match
Best Local Similarity 9.8%; Score 79; DB 2; Length 365;
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;

Db 135 ITLPL-FGWAQNVNDKVCILSQDFGTYIS-TAVAFYIPMSVLMFY 181
QY 2 VTIPPLVFRWFP-VGFRVCKISSPFAFTTPRPHNEVYIGFPITLLHF 49

RESULT 8
ID US-08-428-243-9 STANDARD; PRT; 365 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 7, Application US/08428243
Sequence 7, Application US/08428243
GENERAL INFORMATION:
APPLICANT: The United States of America, as represented by
APPLICANT: the Secretary of Health and Human Services
TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR
NUMBER OF SEQUENCES: 9
```

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR
CC STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR
CC CITY: NEWPORT BEACH
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 92660
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/428,243
CC FILING DATE: 18-SEP-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/10301
CC FILING DATE: 27-OCT-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Altman, Daniel E.
CC REGISTRATION NUMBER: 34,115
CC REFERENCE/DOCKET NUMBER: NIH046.001VPC
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 714-760-0404
CC
CC QUERY MATCH 9.8%; Score 79; DB 2; Length 404;
CC Best Local Similarity 32.7%; Pred. No. 3.21e+01;
CC Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;
DB 172 ITLPLP-FGWAQNVNDKVLISQDFGTYYS-TAVAFYIPMSVLMFY 218
QY 2 VTIPPLVRFEP-VCGFRVCKISSPFAFTPRWPHNEVYIGFPTLLHF 49
RESULT 10
ID PCT-US93-10301-7 STANDARD; PRT; 404 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 7, Application PC/TUS9310301
CC Sequence 7, Application PC/TUS9310301
CC GENERAL INFORMATION:
CC APPLICANT: The United States of America, as represented by
CC APPLICANT: the Secretary of Health and Human Services
CC TITLE OF INVENTION: THE PCT-65 SEROTONIN RECEPTOR
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: KNOBBE, MARTENS, OLSON AND BEAR
CC STREET: 620 NEWPORT CENTER DRIVE 16TH FLOOR
CC CITY: NEWPORT BEACH
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 92660
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US93/10301
CC FILING DATE:

CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Altman, Daniel E.
CC REGISTRATION NUMBER: 34,115
CC REFERENCE/DOCKET NUMBER: NIH046.001VPC
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 714-760-0404
CC TELEFAX: 714-760-9502
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 404 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 404 AA; 45041 MW; 847946 CN;
Query Match 9.8%; Score 79; DB 3; Length 404;
Best Local Similarity 32.7%; Pred. No. 3.21e+01;
Matches 16; Conservative 9; Mismatches 21; Indels 3; Gaps 3;
DB 172 ITLPLP-FGWAQNVNDKVLISQDFGTYYS-TAVAFYIPMSVLMFY 218
QY 2 VTIPPLVRFEP-VCGFRVCKISSPFAFTPRWPHNEVYIGFPTLLHF 49
RESULT 11
ID US-07-996-772A-12 STANDARD; PRT; 422 AA.
XX
AC xxxxxx
XX
DT
DE Sequence 12, Application US/07996772A
CC Sequence 12, Application US/07996772A
CC Patent No. 5472866
CC GENERAL INFORMATION:
CC APPLICANT: Gerald, Christophe
CC APPLICANT: Hartig, Paul R.
CC APPLICANT: Branche, Theresa A.
CC APPLICANT: Weinshank, Richard L.
CC TITLE OF INVENTION: DNA ENCODING 5-HT4A SEROTONIN
CC TITLE OF INVENTION: RECEPTORS AND USES THEREOF
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: COOPER & DUNHAM
CC STREET: 30 ROCKEFELLER PLAZA
CC CITY: NEW YORK
CC STATE: NEW YORK
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/996,772A
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, P., John
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 42667/JPW/TEP
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 977-9550
CC TELEFAX: (212) 664-0525
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 422 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

```
>US-09-209-961-9
Title:
Description: (1-104) from US09209961.pep
Perfect Score: 807
Sequence: 1 MWTPPLVFRWFPVCGFRVC.....HSSRQVTPILSRSGSTFNK 104
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query			DB	ID	Description	Pred. No.
	No.	Score	Match Length				
1	104	12.9	389	2	S70114	probable membrane pro	5.39e-03
2	92	11.4	367	2	S74900	histidinol-phosphate	2.94e-01
3	91	11.3	1456	2	T01397	l-tr transposon polypr	4.05e-01
4	90	11.2	82	2	S5460	hypothetical protein	5.57e-01
5	89	11.0	267	2	D58933	hypothetical protein	7.64e-01
6	89	11.0	700	2	D70951	probable UvrD - Mycob	7.64e-01
7	89	11.0	793	2	JP5339	Smoothed protein pr	7.64e-01
8	88	10.9	280	1	VFX5DS	outer capsid protein	1.05e+00
9	88	10.9	469	2	T17813	crtJ protein - Rhodob	1.05e+00
10	88	10.9	775	1	VPRXHK	outer layer protein V	1.05e+00
11	88	10.9	775	1	VPRXW5	outer layer protein V	1.05e+00
12	87	10.8	631	2	T02472	hypothetical protein	1.43e+00
13	86	10.7	421	2	S35301	rfbH protein - Yersin	1.94e+00
14	86	10.7	775	1	VPRXWM	outer layer protein V	1.94e+00
15	86	10.7	775	1	VPRXSL	outer layer protein V	1.94e+00
16	86	10.7	775	1	VPRXW4	outer layer protein V	1.94e+00
17	86	10.7	775	2	S52165	outer capsid protein	1.94e+00
18	86	10.7	775	1	VPRXPG	outer layer protein V	1.94e+00
19	85	10.5	147	2	S50775	hypothetical protein	2.64e+00
20	85	10.5	234	2	S70539	response regulator ci	2.64e+00
21	85	10.5	280	1	VPRXWA	outer capsid protein	2.64e+00
22	85	10.5	280	1	VPRX16	outer capsid protein	2.64e+00
23	85	10.5	288	2	S75502	hypothetical protein	2.64e+00

```

ENTRY      S52460      #type complete
TITLE      hypothetical protein 109 - Coxiella burnetii
ORGANISM   #formal_name Coxiella burnetii
DATE       08-May-1995 #sequence_revision 21-Jul-1995 #text_change
          09-Sep-1997

ACCESSIONS S52460
REFERENCE   S52460
#authors   Willens, H.; Thiele, D.; Valkova, D.
#submission submitted to the EMBL Data Library, February 1995
#accession S52460
#status    #preliminary
#molecule_type DNA
##residues 1-82 ##label WIL
##cross-references EMBL:X84722; NID:g682754; PID:g682755
SUMMARY    #length 82 #molecular-weight 9545 #checksum 1991

Query Match      11.2%; Score 90; DB 2; Length 82;
Best Local Similarity 20.8%; Pred. No. 5.57e-01;
Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;

Db 28 LLLVPSLVLFMFPISSIDAYLIYQIGILAIWMLCVDLYTYLLASL 75
      :::lllllllll::: :::::ll:::ll:::ll:::ll:::ll:::ll
QY 1 MVTIPLV-ERWEPVCGFRVCKISSPFAFTTPRWPHNEVIGIFITLL 47

RESULT      5
ENTRY      D58933      #type complete
TITLE      hypothetical protein 267 - Cyanidioschyzon merolae
ORGANISM   #formal_name mitochondrion Cyanidioschyzon merolae
DATE       01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change
          01-Feb-1999

ACCESSIONS D58933
REFERENCE   D58933
#authors   Ohta, N.; Sato, N.; Kuroiwa, T.
#journal   Nucleic Acids Res. (1998) 26:5190-5198
#title     Structure and organization of the mitochondrial genome of the
          unicellular red alga Cyanidioschyzon merolae deduced from
          the complete nucleotide sequence.
#cross-references MUID:99030526
#accession D58933
##status   preliminary
##molecule_type DNA
##residues 1-267 ##label ARN
##cross-references GB:D89861; NID:g4115781; PID:d1037524; PID:g4115800
GENETICS   ORF267
#gene      mitochondrion
#genome    mitochondrion
KEYWORDS    #length 267 #molecular-weight 32317 #checksum 5048
SUMMARY

Query Match      11.0%; Score 89; DB 2; Length 267;
Best Local Similarity 37.1%; Pred. No. 7.64e-01;
Matches 13; Conservative 10; Mismatches 9; Indels 3; Gaps 3;

Db 167 NVYI-FPLSLVFESEYRTIKL-AKNY-RRYLLLL 198
      ::lllllllll:::ll:::ll:::ll:::ll:::ll:::ll
QY 37 EVYIGFITLLHFPAPHFQKFSQPAEIFDKRYRVLL 71

RESULT      6
ENTRY      D70951      #type complete
TITLE      probable UvrD - Mycobacterium tuberculosis (strain H37Rv)
ORGANISM   #formal_name Mycobacterium tuberculosis
DATE       17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
          17-Jul-1998

ACCESSIONS D70951
REFERENCE   A70500
#authors   Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
          C.; Harris, D.; Gordon, S.V.; Eiglmeyer, K.; Gas, S.; Barry
          III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
          Chillingworth, T.; Connor, R.; Davies, R.; Devlin, K.;
          Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
          Johnson, A.; Jones, K.; Maitland, N.; Miller, R.;
          Orkin, N.; Pearson, T.A.; Raza, S.; Rutter, D.;
          Seeger, A.P.; Sharp, R.; Skerrett, M.; Smith, R.;
          Spratt, B.G.; Staden, R.; White, O.; Young, P.;
          Ziegler, W.
          #journal   Nature (1998) 393:269-274
          #title     Whole-genome sequence of Mycobacterium tuberculosis
          H37Rv
          #cross-references EMBL:X85175; NID:g682755; PID:g682755
          #accession D70951
          #molecule_type DNA
          #status    preliminary
          #cross-references EMBL:X85175; NID:g682755; PID:g682755
          #checksum 1991

```

```

Hornsby, T.; Jagels, K.; Krogh, A.; McLean, J.; Moule, S.;
Murphy, L.; Oliver, S.; Osborne, J.; Quail, M.A.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.;
Skellton, S.; Squares, S.; Squires, R.; Sulston, J.E.;
Taylor, K.; Whitehead, S.; Barrell, B.G.
#journal      Nature (1998) 393:537-544
#title       Deciphering the biology of Mycobacterium tuberculosis from
              the complete genome sequence.
#cross-references MUID:98295987
#accession    D70951
#status      preliminary; nucleic acid sequence not shown;
              translation not shown
##molecule_type DNA
##residues    1-700 #label COL
##cross-references GB:AL021646; GB:AL123456; NID:G3242278; PID:el248789;
              PID:G2827608
##experimental_source strain H37Rv

GENETICS
SUMMARY      #length 700 #molecular-weight 75603 #checksum 9430
              11.0%; Score 89; DB 2; Length 700;
Query Match Best Local Similarity 34.1%; Pred. No. 7.64e-01;
Matches 15; Conservative 12; Mismatches 16; Indels 1; Gaps 1;

Db 202 LDFDPLLHTAAIENDAVAEEFQDRYCFVVDYQDVTPLQ 245
QY 40 IGFTILLFPAHFQKFSQPAEIFDKRYVLLCNGHQN-PALQ 82
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

RESULT 7
ENTRY   #type complete
TITLE   Smoothened protein precursor - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE    02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change
17-Mar-1999
ACCESSIONS JC5539; PC4476
REFERENCE   JC5539
AUTHORS    Akiyama, H.; Shigeno, C.; Hiraki, Y.; Shukunami, C.; Kohno,
              H.; Akagi, M.; Konishi, J.; Nakamura, T.
JOURNAL    Biochem. Biophys. Res. Commun. (1997) 235:142-147
TITLE      Cloning of a mouse Smoothed cDNA and expression patterns of
              hedgehog signalling molecules during chondrogenesis and
              cartilage differentiation in clonal mouse EC cells, ATDC5.
#cross-references MUID:97339452
#accession JC5539
##molecule_type mRNA
##residues 1-793 #label AKI
#accession PC4476
##molecule_type protein
##residues 528-533:539-545;600-605 #label AK2
##experimental_source ATDC5 cell
COMMENT    This protein is used in the conserved targets in Hedgehog
              signalling pathway, together with Patched and Gli. These protein
              are responsible for the skeletal abnormalities in Gorlin and
              Greig syndromes.
FEATURE    #domain signal sequence #status predicted #label SIG
1-32       #length 793 #molecular-weight 87299; #checksum 8609
SUMMARY    Query Match 11.0%; Score 89; DB 2; Length 793;
              Best Local Similarity 27.9%; Pred. No. 7.64e-01;
              Matches 12; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 461 FLAFGFLVITSCHFYDFNOAE-WERSFRDYVLCQAVNTIGL 502
QY 39 YIGFPITLLFPAHFQKFSQPAEIFDKRYV-LLCNGHQN-PAL 80
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

RESULT 8
ENTRY   #type complete
TITLE   outer capsid protein VP8 - human rotavirus A (serotype 2
              strain DSI)

```

```

ORGANISM    #formal_name human rotavirus A
DATE        31-Mar-1989 #sequence_revision 31-Mar-1989 #text_change
30-Sep-1993
ACCESSIONS  C25904
REFERENCE    A94126
AUTHORS     Goriglia, M.; Hoshino, Y.; Buckler-White, A.; Blumentals,
              I.; Glass, R.; Flores, J.; Kapikian, A.Z.; Chanock, R.M.
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. (1986) 83:7039-7043
TITLE       Conservation of amino acid sequence of VP8 and cleavage
              region of 84-kDa outer capsid protein among rotaviruses
              recovered from asymptomatic neonatal infection.
#cross-references MUID:86313706
#accession  C25904
##molecule_type genomic RNA
##residues 1-280 #label GOR
COMMENT     VP8 is one of two trypsin cleavage products of VP3; the other
              product is VP5.
GENETICS    #map_position segment 4
              CLASSIFICATION #superfamily rotavirus outer layer protein VP3
              KEYWORDS     glycoprotein; outer capsid protein
FEATURE     56,132,150,195 #binding_site carbohydrate (Asn) (covalent) #status
              predicted
SUMMARY     #length 280 #molecular-weight 32010 #checksum 1604
              10.9%; Score 88; DB 1; Length 280;
Query Match Best Local Similarity 39.4%; Pred. No. 1.05e+00;
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINGLPP-IQNRNVVPLSLSSRSIQYKR 246
QY 72 CNGHNPALQOQTHSSRQVTPLSLRSSTFNK 104
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

RESULT 5
ENTRY   #type complete
TITLE   crtJ protein - Rhodobacter capsulatus
ORGANISM #formal_name Rhodobacter capsulatus
DATE    30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change
30-Jun-1992
ACCESSIONS SI7813
REFERENCE   SI7803
AUTHORS     Burke, D.H.; Alberti, M.; Armstrong, G.A.; Hearst, J.E.
JOURNAL     Submitted to the EMBL Data Library November 1991
TITLE       The complete nucleotide sequence of the 45 kb photosynthesis
              gene cluster of Rhodobacter capsulatus.
#accession SI7813
##molecule_type DNA
##residues 1-469 #label EMB
##cross-references EMBL:Z11165
GENETICS
SUMMARY    crtJ
              #length 469 #molecular-weight 51331 #checksum 8204
              10.9%; Score 88; DB 2; Length 469;
Query Match Best Local Similarity 35.0%; Pred. No. 1.05e+00;
Matches 14; Conservative 10; Mismatches 13; Indels 3; Gaps 3;

Db 147 YEAQREI-ETRYRVLL-EAHPAPLLIVSMSTGR-IADLNL 183
QY 56 FSQPAEIFDKRYVLLCNGHQNPAQOQTHSSRQVTPLSL 95
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |
      :| | | | | :| | | | | :| | | | | :| | | | | :| | | | |

RESULT 10
ENTRY   #type complete
TITLE   outer layer protein VP3 - human rotavirus A (strain KU)
ALTERNATE_NAMES glycoprotein VP3; hemagglutinin; outer capsid protein VP3
ORGANISM #formal_name human rotavirus A
DATE    30-Sep-1989 #sequence_revision 30-Sep-1989 #text_change
21-Nov-1997
ACCESSIONS A28844
REFERENCE   A28844

```

```

RESULT 12
ENTRY T02472 #type complete
TITLE hypothesized protein F4I18.27 - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear
          : cress
DATE 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change
ACCESSIONS T02472
REFERENCE Z14674
#authors Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.;
         Brandon, R.C.; Sykes, S.M.; Kaul, S.; Mason, T.M.;
         Kerlavage, A.R.; Adams, M.D.; Somerville, C.R.; Venter,
         J.C.
#submission submitted to the EMBL Data Library, August 1998
#description Arabidopsis thaliana chromosome II BAC F4I18 genomic
             sequence.
#accesion T02472
##status preliminary; translated from GB/EMBL/DDBJ
##molecule_type DNA
##residues 1-631 #label ROU
##cross-references EMBL:AC004665; NID:g3386593; PID:g3386620
GENETICS
#map_position 2
#introns 218/3; 289/1; 389/2; 523/2
#note F4I18.27
SUMMARY #length 631 #molecular-weight 72101 #checksum 5374

Query Match 10.8%; Score 87; DB 2; Length 631;
Best Local Similarity 50.0%; Pred. No. 1.43e+00;
Matches 8; Conservative 6; Mismatches 1; Indels 1; Gaps 1;

Db 130 CRIPADYGYKTPRWP 145
QY 20 KRISPPAFTTP-RWP 34
   ::::: ||||
   |::|::: ||||

RESULT 13
ENTRY S35301 #type complete
TITLE rfbH protein - Yersinia enterocolitica
ORGANISM #formal_name Yersinia enterocolitica
DATE 13-Jan-1995 #sequence_revision 06-Jun-1997 #text_change
        17-Mar-1999
ACCESSIONS S35301; S28584
REFERENCE S35292
#authors Zhang, L.; Al-Hendy, A.; Tolvanen, P.; Skurnik, M.
#journal Mol. Microbiol. (1993) 9:309-321
#title Genetic organization and sequence of the rfb gene cluster of
       Yersinia enterocolitica serotype O:3; similarities to the
       gTDP-L-rhamnose biosynthesis pathway of Salmonella and to
       the bacterial polysaccharide transport systems.
#cross-references M01D:94018626
#accesion S35301
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-421 #label ZHA
##cross-references EMBL:Z18920; NID:g48582; PID:g48592
##experimental_source serotype O:3
##note the nucleotide sequence was submitted to the EMBL Data
      Library, December 1992
GENETICS
SUMMARY rfbH
        #length 421 #molecular-weight 47718 #checksum 3631

Query Match 10.7%; Score 86; DB 2; Length 421;
Best Local Similarity 35.0%; Pred. No. 1.94e+00;
Matches 14; Conservative 7; Mismatches 17; Indels 2; Gaps 2;

Db 86 TTITWTFVDS-N-VNMGGHYTLRFANHFATRNIIY 123
QY 24 SPFAFTTPRHNVIGFPITLLHPAHFQNFSPAEIF 63

```

```

RESULT      14
ENTRY       VPXRWM          #type complete
TITLE       outer layer protein VP3 - human rotavirus A (strain L27)
CONTAINS    VP4 protein
            outer capsid protein VP5; outer capsid protein VP8
ORGANISM    #formal_name human rotavirus A
DATE        31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
            01-Mar-1996
ACCESSIONS  C36410
REFERENCE   A36410
AUTHORS     Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.;
            Urasawa, S.
JOURNAL     J. Virol. (1990) 64:5640-5644
TITLE       Nucleotide sequence of VP4 and VP7 genes of human rotaviruses
            with subgroup I specificity and long RNA pattern:
            implication for new G serotype specificity.
CROSS-REFERENCES MUID:91012813
ACCESSION   C36410
MOLECULE_TYPE genomic RNA
RESIDUES    1-775 #label TAN
CROSS-REFERENCES EMBL:M58292
GENETICS
MAP_POSITION segment 4
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
KEYWORDS      capsid protein; glycoprotein
FEATURE
1-240         #product outer capsid protein VP8 #status predicted
              #label.VP8\
241-246       #region cleavage processing #status predicted\
247-775       #product outer capsid protein VP5 #status predicted
              #label.VP5\
32,56,97,132,150.
195,324,583,589.
599           #binding_site carbohydrate (Asn) (covalent) #status...
              predicted
SUMMARY       #length.775 #molecular-weight 87506 #checksum 5913
Query Match 10.7%; Score 86; DB 1; Length 775;
Best Local Similarity 39.4%; Pred. NO. 1.94e+00;
Matches 13; Conservative 6; Mismatches 13; Indels 1; Gaps 1;

DbB 215 CNEYINNGLP-IQTRNVPLSLSRSIQYRR 246
|| | : ||| : ||| ||| ||| :
QY 72 CNQHONPALQQGTHSRQVTPLSLRSSSTFNK 104

RESULT      15
ENTRY       VPXRWL          #type complete
TITLE       outer layer protein, VP3|- human rotavirus A (strain L26)
CONTAINS    VP4 protein
            outer capsid protein VP5; outer capsid protein VP8
ORGANISM    #formal_name human rotavirus A
DATE        31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change
            05-Sep-1997
ACCESSIONS  A36410
REFERENCE   A36410
AUTHORS     Taniguchi, K.; Urasawa, T.; Kobayashi, N.; Gorziglia, M.;
            Urasawa, S.
JOURNAL     J. Virol. (1990) 64:5640-5644
TITLE       Nucleotide sequence of VP4 and VP7 genes of human rotaviruses
            with subgroup I specificity and long RNA pattern:
            implication for new G serotype specificity.
CROSS-REFERENCES MUID:91012813
ACCESSION   A36410
MOLECULE_TYPE genomic RNA
RESIDUES    1-775 #label TAN
CROSS-REFERENCES EMBL:M58292; NID:g333854; PID:g333855
GENETICS
MAP_POSITION segment 4
CLASSIFICATION #superfamily rotavirus outer layer protein VP3
KEYWORDS      capsid protein; glycoprotein
FEATURE

```

MAISREH

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:23:20 1999; MasPar time 5.92 Seconds
Tabular output not generated.
--496.197 Million cell updates/sec

Title: >US-09-209-961-9
Description: (1-104) from US09209961.pep
Perfect score: 807
Sequence: 1 MVTIPPLVFRWFPVCGFRVC.....HSSQVTPLSLRSTFNK 104

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 40.593; Variance 66.688; scale 0.609

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	92	11.4	367	1 HIS8_SVNY3	HISTIDINOL-PHOSPHATE A	7.25e-02
2	88	10.9	469	1 CRTJ_RHOCA	CRTJ PROTEIN.	2.97e-01
3	88	10.9	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	2.97e-01
4	88	10.9	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	2.97e-01
5	87	10.8	524	1 CP51_UNCNE	CYTOTOXIC P450 51 (EC 4.20e-01	5.92e-01
6	86	10.7	775	1 VP4_ROTTP	OUTER CAPSID PROTEIN V	5.92e-01
7	86	10.7	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	5.92e-01
8	86	10.7	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	5.92e-01
9	85	10.5	147	1 YJVL_YEAST	VERY HYPOTHEICAL 16.8	8.32e-01
10	85	10.5	234	1 CITE_KLEPN	TRANSCRIPTIONAL REGULA	8.32e-01
11	85	10.5	355	1 NIFS_LACDE	NIFS PROTEIN HOMOLOG (8.32e-01
12	85	10.5	440	1 LCAT_RAT	PHOSPHATIDYLCHOLINE-ST	8.32e-01
13	85	10.5	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	8.32e-01
14	85	10.5	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	8.32e-01
15	85	10.5	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	8.32e-01
16	84	10.4	153	1 ELAA_ECOLI	ELAA PROTEIN.	1.17e+00
17	84	10.4	201	1 GT_ECOLI	GLUTATHIONE S-TRANSFER	1.17e+00
18	84	10.4	440	1 LCAT_HUMAN	PHOSPHATIDYLCHOLINE-ST	1.17e+00
19	84	10.4	457	1 UVRD_MCLE	PUTATIVE DNA HELICASE	1.17e+00
20	82	10.2	352	1 YDGO_ECOLI	HYPOTHEICAL 38.1 KD P	2.27e+00
21	82	10.2	438	1 LCAT_MOUSE	PHOSPHATIDYLCHOLINE-ST	2.27e+00
22	82	10.2	440	1 LCAT_PAPAN	PHOSPHATIDYLCHOLINE-ST	2.27e+00
23	82	10.2	775	1 VP4_ROTTH	OUTER CAPSID PROTEIN V	2.27e+00

ID	HIS8_SVNY3	STANDARD;	PRT;	367 AA.
AC	P73807			
DT	15-JUL-1998	(REL. 36, CREATED)		
DT	15-JUL-1998	(REL. 36, LAST SEQUENCE UPDATE)		
DT	15-JUL-1998	(REL. 36, LAST ANNOTATION UPDATE)		
DE	HISTIDINOL-PHOSPHATE AMINOTRANSFERASE (EC 2.6.1.9) (IMIDAZOLE ACETOL-PHOSPHATE TRANSAMINASE)			
GN	HIS8 OR SLL1958			
JS	SYNECHOCYSTIS SP. (STRAIN PCC 6803)			
OC	BACTERIA: CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
XX	MEDLINE; 97061201.			
XA	KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y., MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T., HOSOUCHI T., MATSUO A., MURAKI N., NAKAZAKI N., NARUO K., OKUMURA S., SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M., TABATA S.			
RA	Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.			
RL	DNA RES. 3:109-136(1996).			
CC	1- CATALYTIC ACTIVITY: L-HISTIDINOL-PHOSPHATE + 2-OXOGLUTARATE -> 3-(IMIDAZOL-4-YL)-2-OXOPROPYL PHOSPHATE + GLUTAMATE.			
CC	1- COFACTOR: PYRIDOXAL PHOSPHATE.			
CC	1- PATHWAY: EIGHTH STEP IN HISTIDINE BIOSYNTHETIC PATHWAY.			
CC	1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	EMBL; D90909; G1652943;			
DR	PROSITE; P500599; AA_TRANSFER_CLASS_2; 1.			
RP	PF00222; aminotran_2; 1			
TW	HISTIDINE BIOSYNTHESIS; TRANSFERASE; AMINOTRANSFERASE;			
KW	PYRIDOXAL PHOSPHATE			
FT	BINDING 225			
FT	SEQUENCE 367 AA; 40686 MW; 05AA51CC CRC32;			
SO	Query Match 11.4%; Score 92; DB 1; Length 367;			

Best Local Similarity 37.58; Pred. No. 7.25e-02; Matches 9; Conservative 7; Mismatches 7; Indels 1; Gaps 1;

Db 297 FOVCPDSANFEVFAAPRWMAADLY 320
QY 17 FRVCKISSPFAITPRW-PHNEVY 39

RESULT 2
ID CRTJ_RHOCA STANDARD; PRT; 469 AA.
AC P26167;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE CRTJ PROTEIN.
GN CRTJ.
OS RHODOBACTER CAPSULATUS (RHODOPSEUDOMONAS CAPSULATA).
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOBACTER GROUP;
OC RHODOBACTER.
RN [1]

SEQUENCE FROM N.A.
RA BURKE D.H., ALBERTI M., ARMSTRONG G.A., HEARST J.E.;
RL SUBMITTED (NOV-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- PATHWAY: CAROTENOID AND CHLOROPHYLL BIOSYNTHESIS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z11165; G995679;
DR PIR; S17813; S17813.
DR PFAM; PF00158; sigma54; 1.
KW PHOTOSYNTHESIS; CHLOROPHYLL BIOSYNTHESIS; CAROTENOID BIOSYNTHESIS.
SQ SEQUENCE 469 AA; 51363 MW; 6C3D4D39 CRC32;

Query Match 10.9%; Score 88; DB 1; Length 469;
Best Local Similarity 35.08; Pred. No. 2.97e-01;
Matches 14; Conservative 10; Mismatches 13; Indels 3; Gaps 3;

Db 147 YEAOEIE-ETRYRVLL-EAHPAPLLIVSMSTGR-IAIDLNL 183
QY 56 FSQPAEIEFDRIYRYLLGNGHQPALQOQTHSSRQVTPLSL 95

RESULT 3
ID VP4_ROTUP STANDARD; PRT; 775 AA.
AC P11195;
DT 01-JUL-1989 (REL. 11, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN S4.
OS HUMAN ROTAVIRUS (SEROTYPE 3 / STRAIN P).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88275070.
RA GORZIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R.,
RA KAPIKIAN A.Z., CHANOCK R.M.;
RT "Sequence of the fourth gene of human rotaviruses recovered from
RT asymptomatic or symptomatic infections."
RL J. VIROL. 62:2978-2984(1988).
RN [2]
RP SEQUENCE OF 1-280.
RX MEDLINE; 86313706.
RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.,
RA FLORES J., KAPIKIAN A.Z., CHANOCK R.M.;
RT "Conservation of amino acid sequence of VP8 and cleavage region of

RT 84-kDa outer capsid protein among rotaviruses recovered from
RT asymptomatic neonatal infection."
RL PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
DR PIR; C28839; VPXRWS.
DR PIR; E25904; VPXRPP.
DR PFAM; PF00426; VP4; 1.
KW COAT PROTEIN; GLYCOPROTEIN.
FT CHAIN 1 775
FT CHAIN 247 775
FT CHAIN 1 240
FT CHAIN 247 775
FT CARBOHYD 32 32
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 129 129
FT CARBOHYD 132 132
FT CARBOHYD 132 132
FT CARBOHYD 195 195
FT CARBOHYD 324 324
FT CARBOHYD 583 583
FT CARBOHYD 589 589
FT CARBOHYD 592 592
FT CARBOHYD 599 599
FT CARBOHYD 599 599
FT CONFLICT 35 35
FT CONFLICT 192 192
FT CONFLICT 236 237
FT CONFLICT 236 237
SQ SEQUENCE 775 AA; 87687 MW; 8A9F10C1 CRC32;

Query Match 10.9%; Score 88; DB 1; Length 775;
Best Local Similarity 39.4%; Pred. No. 2.97e-01;
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINNGLPP-IONTRNVVPLSLSSRSIQYKR 246
QY 72 CNGHQPALQOQTHSSRQVTPLSLRSSTFNK 104

RESULT 4
ID VP4_ROTUK STANDARD; PRT; 775 AA.
AC P13842;
DT 01-JAN-1990 (REL. 13, CREATED)
DT 01-JAN-1990 (REL. 13, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN S4.
OS HUMAN ROTAVIRUS (STRAIN KU).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 88230603.
RA TANIGUCHI K., MALLOY W.L., NISHIKAWA K., GREEN K.Y., HOSHINO Y.,
RA URASAWA S., KAPIKIAN A.Z., CHANOCK R.M., GORZIGLIA M.;
RT "Identification of cross-reactive and serotype 2-specific
RT neutralization epitopes on VP3 of human rotavirus."
RL J. VIROL. 62:2421-2426(1988).
CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -1- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M21014; G333853;
DR PIR; A28844; VPXRHK.

DR PFAM: PF00426; VP4; 1.
KW COAT PROTEIN; GLYCOPROTEIN.
FT CHAIN 1 775
FT CHAIN 1 240
FT CHAIN 247 775
FT CARBOHYD 32 32
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 132 132
FT CARBOHYD 324 324
FT CARBOHYD 583 583
FT CARBOHYD 589 589
FT CARBOHYD 592 592
FT CARBOHYD 599 599
SQ SEQUENCE 775 AA; 87538 MW; AA26C921 CRC32;
Query Match 10.9%; Score 88; DB 1; Length 775;
Best Local Similarity 39.4%; Pred. No. 2.97e-01;
Matches 13; Conservative 7; Mismatches 12; Indels 1; Gaps 1;
Db 215 CNEYINNGLPP-IQONRVNVPVLSRSRSTYQR 246
QY 72 CNGHQNFALQOQTHSSRQVTPLSLRSRSTFNK 104
RESULT 5
ID CP51_UNCNE STANDARD; PRT: 524 AA.
AC 014422; 014422;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE CYTOCHROME P450 51 (EC 1.14.14.1) (CYPL1) (P450-L1A1) (P450-L14-
DE ALPHA DEMETHYLASE) (EBURICOL 14-ALPHA-DEMETHYLASE) (P450-14DM).
GN CYP51.
OS UNCINULA NECATOR (GRAPE POWDERY MILDEW).
OC EURKYOTA; FUNGI; ASCOMYCOTA; EUASCOMYCETES; ERYSIPTHALES;
OC ERYSIPTHACEAE; UNCINULA.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-FPE11;
RX MEDLINE; 97444284.
RA DELYE C., LAIGRET F., CORIO-COSTET M.-F.;
RT "Cloning and sequence analysis of the eburicol 14alpha-demethylase
RT gene of the obligate biotrophic grape powdery mildew fungus.";
RL GENE 195:29-33(1997).
RN [2]
RP MOLECULAR BASIS OF FUNGICIDE RESISTANCE.
RC STRAIN-P4211;
RC MEDLINE; 97394921.
RA DELYE C., LAIGRET F., CORIO-COSTET M.-F.;
RX "A mutation in the 14 alpha-demethylase gene of Uncinula necator that
RX correlates with resistance to a sterol biosynthesis inhibitor.";
RT APPL. ENVIRON. MICROBIOL. 63:2966-2970(1997).
CC -!- FUNCTION: CATALYSES C14-DEMETHYLATION OF LANOSTEROL WHICH IS
CC CRITICAL FOR ERGOSTEROL BIOSYNTHESIS. IT TRANSFORMS LANOSTEROL
CC INTO 4,4'-DIMETHYL CHOLESTA-8,14,24-TRIENE-3-BETA-OL
CC (BY SIMILARITY).
CC -!- PATHWAY: ERGOSTEROL BIOSYNTHESIS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; U72657; G2406574;
CC EMBL; U72658; G2406576;
CC EMBL; U83840; G2327023;
CC PROSITE; PS00086; CYTOCHROME_P450; 1.
CC PFAM; PF00067; p450; 1.

KW ELECTRON TRANSPORT; OXIDOREDUCTASE; MONOOXYGENASE; MEMBRANE; HEME;
KW STEROL BIOSYNTHESIS.
FT BINDING 469 469
FT VARIANT 136 136
FT TRIADIMENOL; A STEROL DEMETHYLATION-
FT INHIBITING FUNGICIDE).
SQ SEQUENCE 524 AA; 59855 MW; 10D3D2A2 CRC32;
Query Match 10.8%; Score 87; DB 1; Length 524;
Best Local Similarity 41.7%; Pred. No. 4.20e-01;
Matches 10; Conservative 5; Mismatches 9; Indels 0; Gaps 0;
Db 52 PPVFWHFFPIIGTISYGDIPYKF 75
QY 5 PPIVERWEPVCGFRVCKISSPFAF 28
RESULT 6
ID VP4_ROTGP STANDARD; PRT: 775 AA.
AC P23045;
DT 01-NOV-1991 (REL. 20, CREATED)
DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
GN S4.
OS PORCINE ROTAVIRUS (STRAIN GOTTFRIED).
OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90080150.
RA GORTIGLIA M., NISHIKAWA K., HOSBINO Y., TANIGUCHI K.;
RT "Similarity of the outer capsid protein VP4 of the Gottfried strain
RT of porcine rotavirus to that of asymptomatic human rotavirus
RT strains.";
RL J. VIROL. 64:414-418(1990).
CC -!- SUBCELLULAR LOCATION: OUTER CAPSID.
CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.
CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC EMBL; N33516; G333316;
CC PIR; A33567; VPXRPG.
CC PFAM; PF00426; VP4; 1.
KW COAT PROTEIN; GLYCOPROTEIN.
FT CHAIN 1 775
FT CHAIN 1 240
FT CHAIN 247 775
FT CARBOHYD 32 32
FT CARBOHYD 56 56
FT CARBOHYD 97 97
FT CARBOHYD 132 132
FT CARBOHYD 324 324
FT CARBOHYD 583 583
SQ SEQUENCE 775 AA; 87098 MW; 9F77E893 CRC32;
Query Match 10.7%; Score 86; DB 1; Length 775;
Best Local Similarity 36.4%; Pred. No. 5.92e-01;
Matches 12; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
Db 215 CNEYINNGLPP-MONTRNVVPVLSRSRSTYQR 246
QY 72 CNGHQNFALQOQTHSSRQVTPLSLRSRSTFNK 104

OS HUMAN ROTAVIRUS. (SEROTYPE 2 / STRAIN DSL).

OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.

[1]

RP SEQUENCE FROM N.A.

RA MEDLINE; 88275070

RX GORZIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R., KAPIKAN A.Z., CHANOCK R.M.;

RT "Sequence of the fourth gene of human rotaviruses recovered from asymptomatic or symptomatic infections.";

ST J. VIROL. 62:2978-2984(1988).

RN [2]

RP SEQUENCE OF 1-280.

RX MEDLINE; 86313706.

RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R., FLORES J., KAPIKAN A.Z., CHANOCK R.M.;

RT "Conservation of amino acid sequence of VP8 and cleavage region of 84-kDa outer-capsid protein among rotaviruses recovered from asymptomatic neonatal infection.";

ST KY. J. ZOOLOGICAL RES. 15:103-107(1988).

RN [3]

RP PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).

RX CC -1- SUBCELLULAR LOCATION: OUTER CAPSID.

RT CC -1- PTH: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE OTHER PRODUCT IS VP5.

ST CC -1- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.

RN [4]

RP PIR; B28839; VPKRW4.

RX PIR; C25904; VPKRDS.

RA PFAM; PF00426; VP4; 1.

RT COAT PROTEIN; GLYCOPROTEIN.

ST

RN

FT CHAIN 1 775 OUTER CAPSID PROTEIN VP4.

FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.

FT CHAIN 247 775 OUTER CAPSID PROTEIN VP5.

FT CARBOHYD 56 56 POTENTIAL.

FT CARBOHYD 132 132 POTENTIAL.

FT CARBOHYD 150 150 POTENTIAL.

FT CARBOHYD 195 195 POTENTIAL.

FT CARBOHYD 324 324 POTENTIAL.

FT CARBOHYD 583 583 POTENTIAL.

FT CARBOHYD 589 589 POTENTIAL.

FT CARBOHYD 592 592 POTENTIAL.

FT CARBOHYD 599 599 POTENTIAL.

FT CONFLICT 52 53 HG -> SW (IN REF. 2).

FT CONFLICT 106 107 IA -> SS (IN REF. 2).

FT CONFLICT 142 144 MFK -> TFT (IN REF. 2).

FT CONFLICT 245 245 R -> K (IN REF. 2).

FT CONFLICT 280 280 V -> I (IN REF. 2).

ST SEQUENCE 775 AA; 87592 MW; C01DD686 CRC32;

QY

Db 215 CNEYINNGLP-IONTRNVPLSLSSRSIQYRR 246

QY 72 CNGHQNALQOGTHSSRQVTPLSLRSSTENK 104

RESULT 9

ID YJVL YEAST STANDARD; PRT; 147 AA.

AC P40898;

DT 01-FEB-1995 (REL. 31, CREATED)

DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)

DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE VERY HYPOTHETICAL 16.8 KD PROTEIN IN CRT1 REGION.

GN YJL211C OR J0238 OR HRD147.

OS EUKARYOTAS CEREVISIAE (BAKER'S YEAST).

OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES; SACCHAROMYCETACEAE; SACCHAROMYCETES.

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN-S288C.

RX MEDLINE; 95242842.

RA VANDENBOL M., DURAND P., BOLLE P.-A., DION C., PORTELLE D., HILGER F.;

RT "Sequence analysis of a 40.2 kb DNA fragment located near the left

telomere of yeast chromosome X.;

Yeast 10:1657-1662(1994).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z34098; G496952; -

EMBL: Z49486; G101593; -

PIR: S45163; S45163.

HYPOTHETICAL PROTEIN.

SEQUENCE 147 AA; 16822 MW; 6038A967 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 147;

Best Local Similarity 55.0%; Pred. No. 8.32e-01;

Matches 11; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Db 79 OFFIKRYVELNCRH-AAL 97

51 E1FDKRYVLLNGHNPAL 80

RESULT 10

ID C1TB_KLEPN STANDARD; PRT: 234 AA.

AC P52688;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRANSCRIPTIONAL REGULATORY PROTEIN C1TB.

GN C1TB.

OS KLEBSIELLA PNEUMONIAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA-SUBDIVISION; ENTEROBACTERIACEAE;

OC KLEBSIELLA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 13882;

EX MEDLINE; 96342382.

RA BOTT M., MEYER M., DIMROTH P.;

RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae."

RL MOL. MICROBIOL. 18:533-546(1995).

CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CITA/CITB ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- PTM: PHOSPHORYLATED BY CITA (POTENTIAL).

CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U31464; G924992; -

PFAM: PF00072; response_reg.1.

KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR; PHOSPHORYLATION;

KW SENSORY TRANSDUCTION.

FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).

FT DNA_BIND 181 200 H-T-H MOTIF (BY SIMILARITY).

SQ SEQUENCE 234 AA; 26820 MW; 03C399F6 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 234;

Best Local Similarity 34.4%; Pred. No. 8.32e-01;

Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 107 PVHYORLQHTLERFA-RYRSLRSSEQASQLH 137

telomere of yeast chromosome X.;

Yeast 10:1657-1662(1994).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z34098; G496952; -

EMBL: Z49486; G101593; -

PIR: S45163; S45163.

HYPOTHETICAL PROTEIN.

SEQUENCE 147 AA; 16822 MW; 6038A967 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 147;

Best Local Similarity 55.0%; Pred. No. 8.32e-01;

Matches 11; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Db 79 OFFIKRYVELNCRH-AAL 97

51 E1FDKRYVLLNGHNPAL 80

RESULT 10

ID C1TB_KLEPN STANDARD; PRT: 234 AA.

AC P52688;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRANSCRIPTIONAL REGULATORY PROTEIN C1TB.

GN C1TB.

OS KLEBSIELLA PNEUMONIAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA-SUBDIVISION; ENTEROBACTERIACEAE;

OC KLEBSIELLA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 13882;

EX MEDLINE; 96342382.

RA BOTT M., MEYER M., DIMROTH P.;

RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae."

RL MOL. MICROBIOL. 18:533-546(1995).

CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CITA/CITB ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- PTM: PHOSPHORYLATED BY CITA (POTENTIAL).

CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U31464; G924992; -

PFAM: PF00072; response_reg.1.

KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR; PHOSPHORYLATION;

KW SENSORY TRANSDUCTION.

FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).

FT DNA_BIND 181 200 H-T-H MOTIF (BY SIMILARITY).

SQ SEQUENCE 234 AA; 26820 MW; 03C399F6 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 234;

Best Local Similarity 34.4%; Pred. No. 8.32e-01;

Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 107 PVHYORLQHTLERFA-RYRSLRSSEQASQLH 137

telomere of yeast chromosome X.;

Yeast 10:1657-1662(1994).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z34098; G496952; -

EMBL: Z49486; G101593; -

PIR: S45163; S45163.

HYPOTHETICAL PROTEIN.

SEQUENCE 147 AA; 16822 MW; 6038A967 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 147;

Best Local Similarity 55.0%; Pred. No. 8.32e-01;

Matches 11; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Db 79 OFFIKRYVELNCRH-AAL 97

51 E1FDKRYVLLNGHNPAL 80

RESULT 10

ID C1TB_KLEPN STANDARD; PRT: 234 AA.

AC P52688;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRANSCRIPTIONAL REGULATORY PROTEIN C1TB.

GN C1TB.

OS KLEBSIELLA PNEUMONIAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA-SUBDIVISION; ENTEROBACTERIACEAE;

OC KLEBSIELLA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 13882;

EX MEDLINE; 96342382.

RA BOTT M., MEYER M., DIMROTH P.;

RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae."

RL MOL. MICROBIOL. 18:533-546(1995).

CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CITA/CITB ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- PTM: PHOSPHORYLATED BY CITA (POTENTIAL).

CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U31464; G924992; -

PFAM: PF00072; response_reg.1.

KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR; PHOSPHORYLATION;

KW SENSORY TRANSDUCTION.

FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).

FT DNA_BIND 181 200 H-T-H MOTIF (BY SIMILARITY).

SQ SEQUENCE 234 AA; 26820 MW; 03C399F6 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 234;

Best Local Similarity 34.4%; Pred. No. 8.32e-01;

Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 107 PVHYORLQHTLERFA-RYRSLRSSEQASQLH 137

telomere of yeast chromosome X.;

Yeast 10:1657-1662(1994).

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: Z34098; G496952; -

EMBL: Z49486; G101593; -

PIR: S45163; S45163.

HYPOTHETICAL PROTEIN.

SEQUENCE 147 AA; 16822 MW; 6038A967 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 147;

Best Local Similarity 55.0%; Pred. No. 8.32e-01;

Matches 11; Conservative 6; Mismatches 2; Indels 1; Gaps 1;

Db 79 OFFIKRYVELNCRH-AAL 97

51 E1FDKRYVLLNGHNPAL 80

RESULT 10

ID C1TB_KLEPN STANDARD; PRT: 234 AA.

AC P52688;

DT 01-OCT-1996 (REL. 34, CREATED)

DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)

DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)

DE TRANSCRIPTIONAL REGULATORY PROTEIN C1TB.

GN C1TB.

OS KLEBSIELLA PNEUMONIAE.

OC BACTERIA; PROTEOBACTERIA; GAMMA-SUBDIVISION; ENTEROBACTERIACEAE;

OC KLEBSIELLA.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 13882;

EX MEDLINE; 96342382.

RA BOTT M., MEYER M., DIMROTH P.;

RT "Regulation of anaerobic citrate metabolism in Klebsiella pneumoniae."

RL MOL. MICROBIOL. 18:533-546(1995).

CC -!- FUNCTION: MEMBER OF THE TWO-COMPONENT REGULATORY SYSTEM CITA/CITB ESSENTIAL FOR EXPRESSION OF CITRATE-SPECIFIC FERMENTATION GENES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- PTM: PHOSPHORYLATED BY CITA (POTENTIAL).

CC -!- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: U31464; G924992; -

PFAM: PF00072; response_reg.1.

KW TRANSCRIPTION REGULATION; DNA-BINDING; ACTIVATOR; PHOSPHORYLATION;

KW SENSORY TRANSDUCTION.

FT MOD_RES 56 56 PHOSPHORYLATION (BY SIMILARITY).

FT DNA_BIND 181 200 H-T-H MOTIF (BY SIMILARITY).

SQ SEQUENCE 234 AA; 26820 MW; 03C399F6 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 234;

Best Local Similarity 34.4%; Pred. No. 8.32e-01;

Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 107 PVHYORLQHTLERFA-RYRSLRSSEQASQLH 137

RA MERONI G., MARGARETTI N., MAGNAGHI P., TAREMELLI R.;
 RT "Nucleotide sequence of the cDNA for lecithin-cholesterol acyl
 RT transferase (LCAT) from the rat."
 RL NUCLEIC ACIDS RES. 18:5308-5308(1990).
 CC -!- FUNCTION: CENTRAL ENZYME IN THE EXTRACELLULAR METABOLISM OF PLASMA
 CC LIPOPROTEINS. AMONG OTHER SUBSTRATES IT ESTERIFIES THE FREE
 CC CHOLESTEROL TRANSPORTED IN PLASMA LIPOPROTEINS.
 CC -!- CATALYTIC ACTIVITY: PHOSPHATIDYLCHOLINE + STEROL - STEROL ESTER +
 CC 1-ACYLGLYCEROPHOSPHOCHOLINE (PALMITOYL, OLEOYL, AND LINOLEOYL CAN
 CC BE TRANSFERRED; A NUMBER OF STEROLS, INCLUDING CHOLESTEROL, CAN
 CC ACT AS ACCEPTOR).
 CC -!- ENZYME REGULATION: APOLIPOPROTEIN A-I IS A POTENT ACTIVATOR FOR
 CC THIS ENZYME.
 CC -!- SIMILARITY: PARTIAL WITH LIPASES (PANCREATIC, GASTRIC, HEPATIC,
 CC LINGUAL, LIPOPROTEIN, BACTERIAL, ETC.).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X54096; G56564; --
 DR PIR: S11214; XARTN.
 DR PROSITE: PS00120; LIPASE_SER; 1.
 DR TRANSFERASE; ACYLTRANSFERASE; LIPID METABOLISM; GLYCOPROTEIN; SIGNAL.
 KW SIGNAL 1 24
 FT CHAIN 25 440 LECITHIN-CHOLESTEROL ACYLTRANSFERASE.
 FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 74 98 BY SIMILARITY.
 FT CARBOHYD 337 380 BY SIMILARITY.
 FT CARBOHYD 44 44 POTENTIAL.
 FT CARBOHYD 108 108 POTENTIAL.
 FT CARBOHYD 296 296 POTENTIAL.
 FT CARBOHYD 397 397 POTENTIAL.
 FT CARBOHYD 408 408 POTENTIAL.
 SQ SEQUENCE 440 AA; 49727 MW; 5DEB1ECD CRC32;

Query Match 10.5%; Score 85; DB 1; Length 440;
 Best Local Similarity 23.2%; Pred. No. 8.32e-01;
 Matches 13; Conservative 18; Mismatches 19; Indels 6; Gaps 6;

Db 268 RI-TTTSMPFPAHVPEDHVFISTP-NF-NYTGQDFEFFADLH-FEGGWHMFL 319
 QY 18 RVKISSPFAFTTPR-WPNEVIGRPTLLHPPAH-FQFSQPAEIFDKRYRVL 71

RESULT 13
 ID VP4_ROTWH STANDARD; PRT; 775 AA.
 AC P11193;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
 DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
 GN S4.
 OS HUMAN ROTAVIRUS (SEROTYPE 1 / STRAIN WA).
 OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88275070.
 RA GORZIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R.,
 RA KAPIKIAN A.Z., CHANOCK R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from
 RT asymptomatic or symptomatic infections."
 RL J. VIROL. 62:2978-2984(1988).
 RN [2]
 RP SEQUENCE OF 1-280.
 RX MEDLINE; 86313706.
 RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.,
 RA KAPIKIAN A.Z., CHANOCK R.M.;
 RT "Conservation of amino acid sequence of VP8 and cleavage region of
 RT 84-kDa outer capsid protein among rotaviruses recovered from
 RT asymptomatic neonatal infections."
 RL J. VIROL. 62:2978-2984(1988).
 RN [2]
 RP SEQUENCE OF 1-280.
 RX MEDLINE; 86313706.
 RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.,
 RA KAPIKIAN A.Z., CHANOCK R.M.;

RT "Conservation of amino acid sequence of VP8 and cleavage region of
 RT 84-kDa outer capsid protein among rotaviruses recovered from
 RT asymptomatic neonatal infection."
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).
 CC -!- SUBCELLULAR LOCATION: OUTER CAPSID.
 CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
 CC OTHER PRODUCT IS VP5.
 CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.
 DR PIR: A28839; VPRXW3.
 DR PIR: D25904; VPRXWA.
 DR PFAM: PF00426; VP4; 1.
 KW COAT PROTEIN; GLYCOPROTEIN.
 FT CHAIN 1 775 OUTER CAPSID PROTEIN VP4.
 FT CHAIN 1 240 OUTER CAPSID PROTEIN VP8.
 FT CARBOHYD 247 775 OUTER CAPSID PROTEIN VP5.
 FT CARBOHYD 32 32 POTENTIAL.
 FT CARBOHYD 56 56 POTENTIAL.
 FT CARBOHYD 97 97 POTENTIAL.
 FT CARBOHYD 129 129 POTENTIAL.
 FT CARBOHYD 132 132 POTENTIAL.
 FT CARBOHYD 195 195 POTENTIAL.
 FT CARBOHYD 324 324 POTENTIAL.
 FT CARBOHYD 385 385 POTENTIAL.
 FT CARBOHYD 577 577 POTENTIAL.
 FT CARBOHYD 583 583 POTENTIAL.
 FT CARBOHYD 589 589 POTENTIAL.
 FT CARBOHYD 592 592 POTENTIAL.
 FT CARBOHYD 599 599 POTENTIAL.
 FT CONFLICT 100 100 D -> E (IN REF. 2).
 FT CONFLICT 265 265 Q -> E (IN REF. 2).
 SQ SEQUENCE 775 AA; 87739 MW; 8B27CF43 CRC32;

Query Match 10.5%; Score 85; DB 1; Length 775;
 Best Local Similarity 36.4%; Pred. No. 8.32e-01;
 Matches 12; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 215 CNEYINCLPP-IQNTNRVPLPLSSRSIQYKR 246
 QY 72 CNGHQNPALQGHSSRQVTPLSLRSRSSTFNK 104

RESULT 14
 ID VP4_ROTWH STANDARD; PRT; 775 AA.
 AC P11194;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-APR-1993 (REL. 25, LAST ANNOTATION UPDATE)
 DE OUTER CAPSID PROTEIN VP4 (HEMAGGLUTININ) (OUTER LAYER PROTEIN VP4)
 DE [CONTAINS: OUTER CAPSID PROTEINS VP5 AND VP8].
 GN S4.
 OS HUMAN ROTAVIRUS (SEROTYPE 4 / STRAIN VA70).
 OC VIRUSES; DSRNA VIRUSES; REOVIRIDAE; ROTAVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 88275070.
 RA GORZIGLIA M., GREEN K., NISHIKAWA K., TANIGUCHI K., JONES R.,
 RA KAPIKIAN A.Z., CHANOCK R.M.;
 RT "Sequence of the fourth gene of human rotaviruses recovered from
 RT asymptomatic or symptomatic infections."
 RL J. VIROL. 62:2978-2984(1988).
 RN [2]
 RP SEQUENCE OF 1-280.
 RX MEDLINE; 86313706.
 RA GORZIGLIA M., HOSHINO Y., BUCKLER-WHITE A., BLUMENTALS I., GLASS R.,
 RA FLORES J., KAPIKIAN A.Z., CHANOCK R.M.;
 RT "Conservation of amino acid sequence of VP8 and cleavage region of
 RT 84-kDa outer capsid protein among rotaviruses recovered from
 RT asymptomatic neonatal infection."
 RL PROC. NATL. ACAD. SCI. U.S.A. 83:7039-7043(1986).
 CC -!- SUBCELLULAR LOCATION: OUTER CAPSID.
 CC -!- PTM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
 CC OTHER PRODUCT IS VP5.
 CC -!- SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.

[WATER] (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:23:46 1999; MasPar time 13.00 Seconds
Tabular output not generated. 436.538 Million cell updates/sec

Title: >US-09-209-961-9
Description: (1-104) from US09209961.pep
Sequence: 1 MVTIPPLVPRWPCGFRVC.....HSSRQVTPLSLRSSTFNK 104

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
9:sp.phage 10:sp.plant 11:sp.rodent 12:sp.unclassified
13:sp.vertebrate 14:sp.virus

Statistics: Mean 39.165; Variance 73.303; scale 0.534
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES					
Result	No.	Score	Query Match Length DB ID	Description	Pred. No.
1	807	100.0	104 14	O92286 P12.1.	1.69e-148
2	761	94.3	104 14	O56124 STRAIN PMWS PCV, COMPL	2.48e-138
3	735	91.1	104 14	O93186 ORF3.	1.34e-132
4	104	12.9	289 3	O05521 SIMILARITY TO NADH-UBI	6.10e-03
5	100	12.4	292 11	O35839 LECITHIN:CHOLESTEROL A	2.42e-02
6	99	12.3	379 5	Q21483 M03C11.1 PROTEIN.	3.39e-02
7	92	11.4	413 3	P78759 FISSION YEAST (FRAGMEN	3.46e-01
8	92	11.4	711 5	O45825 T25E12.11 PROTEIN.	3.46e-01
9	91	11.3	507 8	O20281 MATURASE (FRAGMENT).	4.78e-01
10	90	11.2	82 2	O45911 ORF 109.	6.59e-01
11	90	11.2	103 2	O28733 HYPOTHETICAL 12.1 KD P	6.59e-01
12	90	11.2	499 8	O30353 MATURASE (FRAGMENT).	6.59e-01
13	90	11.2	505 8	O20276 MATURASE (FRAGMENT).	6.59e-01
14	89	11.0	150 14	O56350 GP41 (FRAGMENT).	9.06e-01
15	89	11.0	397 4	P78423 CX3C CHEMOKINE PRECURS	9.06e-01
16	89	11.0	506 8	O78332 MATURASE (FRAGMENT).	9.06e-01
17	89	11.0	507 8	O47164 RIBOSOMAL MATURASE.	9.06e-01
18	89	11.0	511 10	O81939 MATURASE (FRAGMENT).	9.06e-01
19	89	11.0	541 5	O09617 HYPOTHETICAL 61.0 KD P	9.06e-01
20	89	11.0	700 2	O53344 PUTATIVE DNA HELICASE	9.06e-01

21	89	11.0	787 4	Q99835	SMOOTHENED.	9.06e-01
22	89	11.0	793 11	O70246	PUTATIVE NEURONAL CELL	9.06e-01
23	89	11.0	793 11	P97698	SMOOTHENED.	9.06e-01
24	88	10.9	273 14	Q67537	GROUP A ROTAVIRUS ISOL	1.24e+00
25	82	10.9	455 5	Q19442	SIMILARITY WITH VARIOU	1.24e+00
26	88	10.9	775 14	Q86185	VP5 AND VP8.	1.24e+00
27	88	10.9	775 14	Q82012	VP4 PROTEIN.	1.24e+00
28	88	10.9	775 14	Q86111	OUTER CAPSID VP4.	1.24e+00
29	87	10.8	274 5	Q21659	RO3A10.6 PROTEIN (FRAG	1.70e+00
30	87	10.8	291 11	O35502	LECITHIN:CHOLESTEROL A	1.70e+00
31	87	10.8	298 11	O35726	LECITHIN:CHOLESTEROL A	1.70e+00
32	87	10.8	299 11	O35573	LECITHIN:CHOLESTEROL A	1.70e+00
33	87	10.8	511 10	O81935	MATURASE (FRAGMENT).	1.70e+00
34	87	10.8	631 10	O90844	F4118.27 PROTEIN.	1.70e+00
35	96	10.7	259 5	O45911	Y17D7B.3 PROTEIN.	2.33e+00
36	96	10.7	291 11	O35746	LECITHIN:CHOLESTEROL A	2.33e+00
37	86	10.7	421 2	O56506	RFPH.	2.33e+00
38	86	10.7	515 10	O81938	MATURASE (FRAGMENT).	2.33e+00
39	86	10.7	534 10	O95303	PHOSPHATE TRANSPORTER.	2.33e+00
40	85	10.7	775 14	Q82118	RNA FOR VP4.	2.33e+00
41	95	10.5	288 2	P73993	HYPOTHETICAL 33.1 KD P	3.17e+00
42	85	10.5	344 5	O45991	ZK1037.9 PROTEIN.	3.17e+00
43	85	10.5	440 11	O35849	LECITHIN:CHOLESTEROL A	3.17e+00
44	35	10.5	507 8	O62949	MATURASE (FRAGMENT).	3.17e+00
45	95	10.5	803 5	Q22343	T08D10.2 PROTEIN.	3.17e+00

ALIGNMENTS

RESULT 1					
ID	O92286	PRELIMINARY;	PRT;	104 AA.	
AC	O92286;				
DT	01-NOV-1998 (TREMBLREL. 08, CREATED)				
WT	01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE	P12.1.				
OC	PORCINE CIRCOVIRUS.				
NC	VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.				
RF	SEQUENCE FROM N.A.				
IC	STRAIN-412;				
RA	WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;				
RT	"Emergence of a new porcine circovirus."				
RL	SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.				
DR	EMBL: AF085695; G3668368;				
SQ	SEQUENCE 104 AA; 12060 MW; 820A1C5D CRC32;				

Query Match 100.0%; Score 807; DB 14; Length 104;
Best Local Similarity 100.0%; Pred. No. 1.69e-148;
Matches 104; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db	1	MVTIPPLVPRWPCGFRVC	KISSPFAFTTPRPHPNEVIGFPIITLLHFPFAHQKFSQPA 60
QV	1	MVTIPPLVPRWPCGFRVC	KISSPFAFTTPRPHPNEVIGFPIITLLHFPFAHQKFSQPA 60
Db	61	EIFDKRVRVLLCNGHONPALQOQTHSSRQVTPLSLRSSTFNK 104	
QV	63	EIFDKRVRVLLCNGHONPALQOQTHSSRQVTPLSLRSSTFNK 104	

RESULT 2					
ID	O56124	PRELIMINARY;	PRT;	104 AA.	
AC	O56124;				
DT	01-JUN-1998 (TREMBLREL. 06, CREATED)				
WT	01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)				
DT	01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)				
DE	STRAIN PMWS PCV, COMPLETE GENOME.				
OC	PORCINE CIRCOVIRUS.				
NC	VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.				
RF	SEQUENCE FROM N.A.				
IC	STRAIN-PMWS PCV;				
RA	MEDLINE; 98241772.				

RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;
 RT "Nucleotide sequence of porcine circovirus associated with
 RL postweaning multisystemic wasting syndrome in pigs.";
 RJ J. VIROL. 72:5262-5267(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PWNS PCV;
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RX MEDLINE; 98418498
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,
 RY ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
 RT "Characterisation of novel circovirus DNAs associated with wasting
 syndromes in pigs.";
 RJ J. GEN. VIROL. 79:2171-2199(1998).

RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RA MEEHAN B.M.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF027217; G2689648;
 DR EMBL; AF055392; G3598811;
 DR EMBL; AF055391; G3598799;
 SQ SEQUENCE 104 AA; 11879 MW; 68446EFF CRC32;

Query Match 94.3%; Score 761; DB 14; Length 104;
 Best Local Similarity 95.28; Pred. No. 2.48e-138;
 Matches 99; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Db 1 MVTIPLVSWFVPCGFRVCKISSPFAFTTTPRPHNDVYIGLPTLLHFFAHFQKFSQPA 60
 QY 1 MVTIPLVSWFVPCGFRVCKISSPFAFTTTPRPHNDVYIGLPTLLHFFAHFQKFSQPA 60

Db 61 EISDKRYVLLCNHQPALQOQTHSSRQVTPLSLRSSTFNK 104
 QY 61 EIPDKRYVLLCNHQPALQOQTHSSRQVTPLSLRSSTFNK 104

RESULT 3
 ID Q93186 PRELIMINARY; PRT; 104 AA.
 AC Q93186;
 DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ORF3
 OS PORCINE CIRCOVIRUS.

OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RX MEDLINE; 98418498
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,
 RY ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
 RT "Characterisation of novel circovirus DNAs associated with wasting
 syndromes in pigs.";
 RJ J. GEN. VIROL. 79:2171-2199(1998).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RA MEEHAN B.M.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF055394; G3598834;
 DR EMBL; AF055393; G3598823;
 SQ SEQUENCE 104 AA; 11898 MW; 01C3FD5F CRC32;

Query Match 91.1%; Score 735; DB 14; Length 104;
 Best Local Similarity 92.28; Pred. No. 1.34e-132;
 Matches 95; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Db 1 MVTIPLVSWFVPCGFRVCKISSPFAFTTTPRPHNDVYIGLPTLLHFFAHFQKFSQPA 60

QY 1 MVTIPLVSWFVPCGFRVCKISSPFAFTTTPRPHNDVYIGLPTLLHFFAHFQKFSQPA 60
 Db 61 EISDKRYVLLCNHQPALQOQTHSSRQVTPLSLRSSTFNK 103
 QY 61 EIPDKRYVLLCNHQPALQOQTHSSRQVTPLSLRSSTFNK 103

RESULT 4
 ID Q05521 PRELIMINARY; PRT; 289 AA.
 AC Q05521;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE SIMILARITY TO NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4.
 GN D9819.10.
 OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOMYCETES; SACCHAROMYCETALES;
 OC SACCHAROMYCETACEAE; SACCHAROMYCETES.

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S288C (AB972);
 RA FULTON L.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-S288C (AB972);
 RA JOHNSTON M., ANDREWS S., BRINKMAN R., COOPER J., DING H., DU Z.,
 RA FAVELLO A., FULTON L., GATTUNG S., GRECO T., KIRSTEN J., KUCABA T.,
 RA HALLSWORTH K., HAWKINS J., HILLIER L., JIER M., JOHNSON D.,
 RA JOHNSTON L., LANGSTON Y., LATREILLE P., LE T., MARDIS E., MENEZES S.,
 RA MILLER N., NHAN M., PAULEY A., PELUSO D., RIFKEN L., RILES L.,
 RA TAICH A., TREVASKIS E., VIGNATI D., WILCOX L., WOHLDMAN P., VAUDIN M.,
 RA WILSON R., WATERSTON R.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.

RC STRAIN-S288C (AB972);
 RA WATERSTON R.;
 RL SUBMITTED (MAY-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [4]
 RP SEQUENCE FROM N.A.

RC STRAIN-S288C (AB972);
 RA JIA Y., CHERRY J.M.;
 RL SUBMITTED (JUN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U51031; G1332640;
 KW UBIQUINONE.
 SQ SEQUENCE 289 AA; 33514 MW; 1A871897 CRC32;

Query Match 12.9%; Score 104; DB 3; Length 289;
 Best Local Similarity 29.7%; Pred. No. 6.10e-03;
 Matches 11; Conservative 13; Mismatches 12; Indels 1; Gaps 1;

Db 3 RVSEIKTFPFGA-KWLEDFVLLIIMLLNYPVYQ 38
 QY 18 RVCKISSPFAFTTTPRPHNDVYIGLPTLLHFFAHFQ 54

RESULT 5
 ID Q35839 PRELIMINARY; PRT; 292 AA.
 AC Q35839;
 DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE LECITHIN:CHOLESTEROL ACYL TRANSFERASE (FRAGMENT).
 GN LCAT.

OS SCIURUS GRISEUS.
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; RODENTIA;
 OC SCIUROGNATHI; SCIURIDAE; SCIURINAE; SCIURUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ROBINSON M., CATZEFIS F.M., BRIOLAY J., MOUCHIROUD D.;
 RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.

RP SEQUENCE FROM N.A.
RC STRAIN-PE745;
RX MEDLINE; 98162722.
RA YOSHIOKA S., KATO K., NAKAI K., OKAYAMA H., NOJIMA H.;
RT Identification of open reading frames in Schizosaccharomyces pombe
RT CDNAS";
RL DNA RES. 4:363-369(1997).
DE EMBL; D89107; D010466;
FT NON_TER 1
SQ SEQUENCE 413 AA; 48320 MW; BF06DE32 CRC32;

Query Match 11.4%; Score 92; DB 3; Length 413;
Best Local Similarity 38.1%; Pred. No. 3.46e-01;
Matches 16; Conservative 9; Mismatches 13; Indels 4; Gaps 4;

D6 91 FPLEPAPWHLEVEYLLF-IKFLSPD-F-RASKAKSLVDRRF 129
QY 27 APTTPRPH-NEVIIGPIVILLHFFPAHFQFSQPAEIFDKRY 67

RESULT 8
ID O45825 PRELIMINARY; PRT; 711 AA.
AC O45825;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
ST 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DE 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DT T2SE12.11 PROTEIN.
CS T2SE12.11.
GN CAENORHADITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SCERINWETA; RHABDITIA; RHADITIDA;
OC RHABDITINA; RHADITOIDEA; RHADITIDAE; PELODERINAE; CAENORHADITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA MATTHEWS L.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
FX MEDLINE; 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
FA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
XA CRAXTON K., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JTER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTING J., LICZY C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
XA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
EA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
EA THIERY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
elégans";
RL NATURE 368:32-38(1994).
DR EMBL; Z82052; EL349871;
SQ SEQUENCE 711 AA; 83567 MW; 50410A0F CRC32;

Query Match 11.4%; Score 92; DB 5; Length 711;
Best Local Similarity 25.5%; Pred. No. 3.46e-01;
Matches 14; Conservative 13; Mismatches 24; Indels 4; Gaps 4;

D6 490 VCQLEROLYSKKNNHFDHVHYDFDIYLSHTTN-NPFTISQ-FIRKYIPL 542
QY 19 VKISSPFAETTPRWPHNEV-YI-GPITLLHFPAPHQFSQPAIFDKRYRVLL 71

RESULT 9
ID O20281 PRELIMINARY; PRT; 507 AA.
AC O20281;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
ST 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DE 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DT XATURASE (FRAGMENT).
GN MATK.
OC STROPHANTHUS DIVARICATUS.
QY CHLOROPLAST.

OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; GENTIANAE; GENTIANALES; APOCYNACEAE; STROPHANTHUS.
 RN [1]

RP SEQUENCE FROM N.A.
 RA ENDRESS M.E., SENNBAD B., NILSSON S., CIVEYREL L., CHASE M.W.,
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;
 RL OPERA BOT. BELG. 7:59-102(1996).
 DR EMBL; 270186; E32376; -;
 KW CHLOROPLAST.
 FT NON_TER
 SQ SEQUENCE 507 AA; 59810 MW; 6B93832E CRC32;

Query Match 11.3%; Score 91; DB 8; Length 507;
 Best Local Similarity 29.5%; Pred. No. 4.78e-01;
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

Db 209 FLLNFFVCEYESIFVLHKNSSHLRP-TSGV-LLERIYFYGKIERLAETFDQANL 266
 QY 12 FVCGFRVCKISSPFAFTTPRPHNEVIGPITLLHPAHFQKFSQPAEIDKRYRVL 71

Db 267 C 267
 QY 72 C 72

RESULT 10
 ID Q45911 PRELIMINARY; PRT; 82 AA.

AC Q45911;
 DT 01-NOV-1996 (TREMBREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBREL. 01, LAST ANNOTATION UPDATE)
 DE ORF 109.
 OC COXIELLA BURNETII.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; COXIELLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-RUSSIAN ISOLATE Q 1182;
 RA WILLEMS H., THIELE D., VALKOVA D.;
 RL SUBMITTED (FEB-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; X84722; G682755; -;
 SQ SEQUENCE 82 AA; 9545 MW; FE63A40C CRC32;

Query Match 11.2%; Score 90; DB 2; Length 82;
 Best Local Similarity 20.8%; Pred. No. 6.59e-01;
 Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;

Db 28 LLLVPSLVFWFPISSIDAYLIVQYGLAIWMLCVDLYTYLLASL 75
 QY 1 MVTIPPLV-FRWFPVCGFRVCKISSPFAFTTPRPHNEVIGFPITLL 47

RESULT 11
 ID Q52873 PRELIMINARY; PRT; 103 AA.

AC Q52873;
 DT 01-JUN-1998 (TREMBREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 12.1 KD PROTEIN.
 OS COXIELLA BURNETII.
 OC PLASMID QPR.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; COXIELLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PRISCILLA Q177;
 RA LAUTENSCHLAGER S., JAEGER C., WILLEMS H., BALJER G.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; Y15898; E1216848; -;
 KW HYPOTHETICAL PROTEIN; PLASMID.
 SQ SEQUENCE 103 AA; 12085 MW; 52F4B0F9 CRC32;

Query Match 11.2%; Score 90; DB 2; Length 103;
 Best Local Similarity 20.8%; Pred. No. 6.59e-01;

Matches 10; Conservative 18; Mismatches 19; Indels 1; Gaps 1;
 Db 28 LLLVPSLVFWFPISSIDAYLIVQYGLAIWMLCVDLYTYLLASL 75
 QY 1 MVTIPPLV-FRWFPVCGFRVCKISSPFAFTTPRPHNEVIGFPITLL 47

RESULT 12
 ID O03053 PRELIMINARY; PRT; 499 AA.

AC O03053;
 DT 01-JUL-1997 (TREMBREL. 04, CREATED)
 DT 01-JUL-1997 (TREMBREL. 04, LAST SEQUENCE UPDATE)
 DT 01-JUL-1997 (TREMBREL. 04, LAST ANNOTATION UPDATE)
 DE MATURASE (FRAGMENT).
 GN MATK.
 OS PICRALIMA NITIDA.
 OC CHLOROPLAST.
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; GENTIANAE; GENTIANALES; APOCYNACEAE; PICRALIMA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ENDRESS M.E., SENNBAD B., NILSSON S., CIVEYREL L., CHASE M.W.,
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;
 RL OPERA BOT. BELG. 7:59-102(1996).
 DR EMBL; 270179; E328637; -;
 KW CHLOROPLAST.
 FT NON_TER
 SQ SEQUENCE 499 AA; 58904 MW; 53AB1C71 CRC32;

Query Match 11.2%; Score 90; DB 8; Length 499;
 Best Local Similarity 29.5%; Pred. No. 6.59e-01;
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

Db 203 FFLYNYVCEYESIFVLHKNSSHLRS-TSGV-LLERIYFYGKIERLVEFAKDFOVNL 260
 QY 12 FVCGFRVCKISSPFAFTTPRPHNEVIGPITLLHPAHFQKFSQPAEIDKRYRVL 71

Db 261 C 261
 QY 72 C 72

RESULT 13
 ID O20276 PRELIMINARY; PRT; 505 AA.

AC O20276;
 DT 01-JAN-1998 (TREMBREL. 05, CREATED)
 DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
 DT 01-JAN-1998 (TREMBREL. 05, LAST ANNOTATION UPDATE)
 DE MATURASE (FRAGMENT).
 GN MATK.
 OS PICRALIMA NITIDA.
 OC CHLOROPLAST.
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; EUDICOTYLEDONS;
 OC ASTERIDAE; GENTIANAE; GENTIANALES; APOCYNACEAE; PICRALIMA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA ENDRESS M.E., SENNBAD B., NILSSON S., CIVEYREL L., CHASE M.W.,
 RA HUYSMANS S., GRAFSTROEM E., BREMER B.;
 RL OPERA BOT. BELG. 7:59-102(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA CIVEYREL L.;
 RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA CIVEYREL L.;
 RL SUBMITTED (JUL-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; 270179; E329795; -;
 KW CHLOROPLAST.
 FT NON_TER
 SQ SEQUENCE 505 AA; 59552 MW; C87B6754 CRC32;

Query Match 11.2%; Score 90; DB 8; Length 505;
 Best Local Similarity 29.5%; Pred. No. 6.59e-01;
 Matches 18; Conservative 11; Mismatches 30; Indels 2; Gaps 2;

Db 209 FFLNFYCEVESIFVFLHNOSSLRS-TSGV-LLEIRIYFGKIERLVEVPKDFQVNL 266
 QY 12 FVCGFRVCKISSPFAFTTPMPHNEVYIGPITLLHPPAHPKFSQPAEIFDKRYRVL 71
 Db 267 C 267
 QY 72 C 72

RESULT 14
 ID O56350 PRELIMINARY; PRT; 150 AA.
 AC O56350;
 DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
 DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)
 DE GP41 (FRAGMENT).
 GN ENV.
 OS HUMAN IMMUNODEFICIENCY VIRUS TYPE 2 (HIV-2).
 OC VIRUSES; RETROID VIRUSES; RETROVIRIDAE; LENTIVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PO2;
 RA HEREDIA A.;
 RL AIDS RES. HUM. RETROVIRUSES 0:0-0(1998).
 DR EMBL; AF039491; G2828072;
 FT NON_TER 1
 FT NON_TER 150 150
 SQ SEQUENCE 150 AA; 17976 MW; 23720A57 CRC32;

Query Match 11.0%; Score 89; DB 14; Length 150;
 Best Local Similarity 47.6%; Pred. No. 9.06e-01;
 Matches 10; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

Db 42 CAFROVCHTTPWFNSTPKW 62
 QY 15 CGFR-VCKISSPFAF-TTPRW 33

RESULT 15
 ID P78423 PRELIMINARY; PRT; 397 AA.
 AC P78423; O00672;
 DT 01-MAY-1997 (TREMBLREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBLREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE CX3C CHEMOKINE PRECURSOR.
 GN A-152E5.2
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
 OC CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 9717111.
 RA BAZAN J.F., BACON K.B., HARDIMAN G., WANG W., SOO K., ROSSI D.,
 RA GREAVES D.R., ZLOTNIK A., SCHALL T.J.;
 RT "A new class of membrane-bound chemokine with a CX3C motif.";
 RL NATURE 385:640-644(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA ADAMS M.D., LOFTUS B.J., ZHOU L., CROSBY M., FUHRMANN J., MASON T.M.,
 RA BRANDON R., KIM U.J., KERLAVAGE A.R., VENTER J.C.;
 RT "Homo sapiens Chromosome 16 BAC clone CIT987SK-A-152E5.";
 RL SUBMITTED (JUN-1998) TO EMBL/GENEBANK/DBJ DATA BANKS.
 DR EMBL; U91835; G189259;
 DR EMBL; U84487; G1888523;
 DR EMBL; AC004382; G3252821;
 DR PFAM; PF00048; 118; 1.
 KW SIGNAL.
 FT SIGNAL 1 24 POTENTIAL.

FT CHAIN 25 397 CX3C-CHEMOKINE.
 SQ SEQUENCE 397 AA; 42202 MW; C8093D7D CRC32;

Query Match 11.0%; Score 89; DB 4; Length 397;
 Best Local Similarity 26.2%; Pred. No. 9.06e-01;
 Matches 17; Conservative 21; Mismatches 23; Indels 4; Gaps 4;

Db 43 IPVALLIHYQNOASCGRRAIILETRQHLFCADPKQWVKDAMQHLDRQAAALT-RN-G 100
 QY 42 FPIITLL-HFPAHFQKFSQPAEIFDKRYRVLCCGHQNPALQOQT-HSSRQVTPLSLRSRS 99
 Db 101 GFEEK 105
 QY 100 STFKN 104

Search completed: Wed Dec 22 10:24:02 1999
 Job time : 16 secs.

WIREH

(TM)

Release 3.1A John F. Collins, BioComputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:11:23 1999; MasPar time 20.50 Seconds
325.700 Million cell updates/sec

Tabular output not generated.

Title: >US-09-209-961-3

Description: (1-314) from US09209961.pep

Perfect Score: 2358

Sequence: 1 MPSKNGRSGPQPKRWFT.....GGQFVTLSPCPPEPYEINY 314

Scoring table: PAM 150

Gap 11

Searched: 170751 seqs, 21266608 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 34.382; Variance 135.259; scale 0.254

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	197	8.4	286	38	BBTV DNA I clone (7-4	2.21e-08
2	193	8.2	285	38	BBTV DNA II clone (2-	5.24e-08
3	139	5.9	286	38	BBTV DNA II V2 Cl ami	4.05e-03
4	136	5.8	286	31	Banana bunchy top vir	7.38e-03
5	97	4.1	611	19	Periplasmic Beta-N-ac	1.22e+01
6	97	4.1	611	39	Hexosaminidase enzyme	1.22e+01
7	92	3.9	165	34	Streptococcus pneumoniae	2.94e+01
8	91	3.9	820	9	Branching enzyme of r	3.50e+01
9	91	3.9	820	11	Rice starch branching	3.50e+01
10	92	3.9	1822	10	Beta subunit of integ	2.94e+01
11	93	3.9	2150	3	Sequence of viral pro	2.47e+01
12	93	3.9	2150	2	Sequence encoded by R	2.47e+01
13	93	3.9	2150	2	Complete human rhinov	2.47e+01
14	92	3.9	2227	3	Hepatitis A virus HM-	2.94e+01
15	92	3.9	2227	27	Human retinol binding	4.95e+01
16	89	3.8	357	9	Human retinol binding	4.95e+01

17	89	3.8	445	27	W27603	Rat neuropeptide Y re	4.95e+01
18	89	3.8	445	22	W15232	Rat neuropeptide Y-Y5	4.95e+01
19	85	3.8	456	25	W29446	Rattus norvegicus neur	4.95e+01
20	85	3.8	456	29	W37092	Rattus norvegicus hyp	4.95e+01
21	89	3.8	456	18	R95940	Rat Y5 receptor	4.95e+01
22	89	3.8	456	26	W29412	Rat hypothalamic neur	4.95e+01
23	89	3.8	466	22	W15233	Mouse neuropeptide Y-	4.95e+01
24	89	3.8	507	3	P50263	Sequence encoded by p	4.95e+01
25	90	3.8	830	37	W80314	CIITA (class II trans	4.17e+01
26	90	3.8	987	31	W57057	Class II trans activa	4.17e+01
27	90	3.8	1106	37	W80312	CIITA (class II trans	4.17e+01
28	90	3.8	1130	31	W57056	Class II trans activa	4.17e+01
29	90	3.8	1130	14	R7452	Class II transactivat	4.17e+01
30	90	3.8	1130	16	R81569	Class II transactivat	4.17e+01
31	90	3.8	1130	37	W80313	CIITA (class II trans	4.17e+01
32	90	3.8	1207	37	W80311	CIITA (class II trans	4.17e+01
33	90	3.8	1463	6	R29940	Deduced from Lelystad	4.17e+01
34	90	3.8	2227	1	R05697	Attenuated hepatitis	4.17e+01
35	90	3.8	4536	28	W41262	Apolipoprotein B-100	4.17e+01
36	88	3.7	445	22	W15230	Human neuropeptide Y-	5.88e+01
37	88	3.7	445	27	W27604	Human neuropeptide Y	5.88e+01
38	88	3.7	455	18	R95939	Human Y5 receptor	5.88e+01
39	88	3.7	455	29	W37093	Homo sapiens hippocam	5.88e+01
40	88	3.7	455	25	W29447	Human hippocampal neu	5.88e+01
41	88	3.7	455	26	W29413	Human hippocampal neu	5.88e+01
42	88	3.7	533	9	R44617	Human retinol binding	5.88e+01
43	86	3.6	398	31	W37988	Amino acid sequence O	8.27e+01
44	86	3.6	429	26	W40801	Alpha-galactosidase A	8.27e+01
45	86	3.6	429	12	R63234	Human alpha-galactosi	8.27e+01

ALIGNMENTS

RESULT 1

ID W87459 standard; Protein: 286 AA.
AC W87459;
DT 10-FEB-1999 (first entry)
DE BBTV DNA I clone (7-4-2) ORF V2 product DNA I-VI.
KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
NW Musa; phloem damage; aphid; nucleic acid detection.
CS Banana bunchy top virus.
FH Key Location/Qualifiers
FT Misc_difference 92 /note= "encoded by GAA"
FT Misc_difference 153 /note= "encoded by AGA"
FT US5846705 A.
PD 08-DEC-1998.
PR 06-APR-1995: 418071.
PS 06-APR-1995: US-418071.
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
WI Soong C, Wu R, You L;
DR WPI: 99-059037/05.
DR N-PSDB: V71832.
PT Nucleic acids having banana bunchy top virus component sequences -
PT used to design primers for use in polymerase chain reaction
PF detection of the virus
PS Claim 1; columns 23-26; 27pp; English.
CC This represents the ORF V2 product (DNA I-VI) of a banana bunchy top
CC virus (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
CC invention provides nucleic acid sequences associated with BBTV that can
CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences
CC (V71832) to V71833 are used as the basis for the construction of PCR
CC primers, to detect BBTV infection. The PCR technique is used for
CC detecting BBTV in plant tissues (preferably banana, especially Musa
CC species). The virus, one of the most important banana species viruses,
CC causes phloem damage and is transmitted by aphids. PCR detection gives
CC accurate, reliable and specific determination of absence or presence of
CC the virus. 286 AA;
SQ Sequence 286 AA;

Query Match 8.4%; Score 197; DB 38; Length 286;
Query Local Similarity 36.5%; Pred. No. 2.21e-08;

Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;
 Db 4 psllkwftlnysaaarenfslilkeedvhyavvgdevapatgkhlglgylslkrlrlg 63
 QY 13 PHKRWFTLNPNSEDERKKIRELPISL-FDYFVIGEGNEEGRTPHLOGFANFVKQTEN 71
 Db 64 glkky-gsrahwelargdeenskyck-qtllilelglfpvngskrkisemva 116
 QY 72 KRW-WYLGARCHIEKAKGTDOQNKKEYCSKGNLLIECGAPRSOG-QRSIDLSTAVS 124
 RESULT 2
 ID W87460 standard; Protein; 285 AA.
 AC W87460;
 DT 10-FEB-1999 (first entry)
 DE BBTV DNA II V2 (D2) ORF V2 product DNA II V2(D2).
 KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
 MU Musa; phloem damage; aphid; nucleic acid detection.
 OS Banana bunchy top virus.
 FH Key Location/Qualifiers
 FT Misc_difference 49 /note- "encoded by CAG"
 FT US5846705-A.
 PN 08-DEC-1998.
 PD 06-APR-1995; 418071.
 PF 06-APR-1995; US-418071.
 PR (BIOT-) DEV CENT BIOTECHNOLOGY.
 PA Soong T, Wu R, You L;
 PI WPI; 99-059037/05.
 DR N-PSDB; V71833.
 DT Nucleic acids having banana bunchy top virus component sequences -
 PT used to design primers for use in polymerase chain reaction
 PT Claim 12: Columns 25-28; 27pp; English.
 CC This represents the ORF V2 product [DNA II V2(D2)] of a banana bunchy top
 CC virus (BBTV) DNA II (clone 2-17) circular single stranded (css) DNA. The
 CC invention provides nucleic acid sequences associated with BBTV that can
 CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences
 CC (V71830 to V71833) are used as the basis for the construction of PCR
 CC primers, to detect BBTV infection. The PCR technique is used for
 CC detecting BBTV in plant tissues (preferably banana, especially Musa
 CC species). The virus, one of the most important banana species viruses,
 CC causes phloem damage and is transmitted by aphids. PCR detection gives
 CC accurate, reliable and specific determination of absence or presence of
 CC the virus.
 SQ Sequence 285 AA;
 Query Match 8.28; Score 193; DB 38; Length 285;
 Best Local Similarity 37.99; Pred. No. 5.24e-08;
 Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;
 Db 5 kwftlnysaaaredflallkeedvhyavvgdevapatgkhlglgylslkrlrlg 64
 QY 16 RWFTLNPNSEDERKKIRELPISL-FDYFVIGEGNEEGRTPHLOGFANFVKQTEN 74
 Db 65 kky-gsrahwelargdeenskyck-qtllilelglfpvngskrkisemva 105
 QY 75 WYLGARCHIEKAKGTDOQNKKEYCSKGNLLIECGAPRSOG 116
 RESULT 3
 ID W87461 standard; Protein; 286 AA.
 AC W87461;
 DT 10-FEB-1999 (first entry)
 DE BBTV DNA II V2 C1 amino acid sequence.
 KW Banana bunchy top virus; BBTV; PCR technique; plant; tissue; banana;
 MU Musa; phloem damage; aphid; nucleic acid detection.
 OS Banana bunchy top virus.
 PN US5846705-A.
 PD 08-DEC-1998.
 PF 06-APR-1995; 418071.
 PR 06-APR-1995; US-418071.
 PA (BIOT-) DEV CENT BIOTECHNOLOGY.

PI Soong T, Wu R, You L;
 DR WPI; 99-059037/05;
 PT Nucleic acids having banana bunchy top virus component sequences -
 PT used to design primers for use in polymerase chain reaction
 PT detection of the virus
 PS Disclosure: Columns 27-30; 27pp; English.
 CC This represents the amino acid sequence of a banana bunchy top virus
 CC (BBTV) DNA II V2 C1 protein. The invention provides nucleic acid
 CC sequences associated with BBTV that can be used in a PCR technique for
 CC detecting BBTV. The nucleic acid sequences (V71830 to V71833) are used as
 CC the basis for the construction of PCR primers, to detect BBTV infection.
 CC The PCR technique is used for detecting BBTV in plant tissues (preferably
 CC banana, especially Musa species). The virus, one of the most important
 CC banana species viruses, causes phloem damage and is transmitted by
 CC aphids. PCR detection gives accurate, reliable and specific determination
 CC of absence or presence of the virus.
 SQ Sequence 286 AA;
 Query Match 5.98; Score 139; DB 38; Length 286;
 Best Local Similarity 29.88; Pred. No. 4.05e-03;
 Matches 25; Conservative 28; Mismatches 23; Indels 8; Gaps 8;
 Db 8 wmtinnpt-tlpv-mrd-eik-ykvqv-drg-qeg-trhvgvgvkmkrsslkqmrfg 60
 QY 17 WFTLNPNSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLOGFANFVKQTEN 76
 Db 61 fpga-hlekrksgsearsymcke 83
 QY 77 LGARCHIEKAKGTDOQNKKEYCSKE 100
 RESULT 4
 ID W54071 standard; Protein; 286 AA.
 AC W54071;
 DT 11-AUG-1998 (first entry)
 DE Banana bunchy top virus Component 1 protein.
 KW BBTV; probe; diagnostic primer; component 1.
 OS Banana bunchy top virus.
 PN US5756708-A.
 PD 26-MAY-1998.
 PF 24-FEB-1994; 202186.
 PR 24-FEB-1994; US-202186.
 PA (UYQO-) UNIV QUEENSLAND TECHNOLOGY.
 PI Burns TM, Dale JL, Harding RM, Karan M;
 DR WPI; 98-321636/28.
 DT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
 PT and primers and for producing virus-resistant plants
 PS Example 1; Fig 11; 59pp; English.
 CC This sequence represents the protein encoded by the component 1 of
 CC the banana bunchy top virus (BBTV). The DNA molecules can be used as
 CC diagnostic probes or primers or can be inserted into plants or other
 CC organisms e.g. to produce virus-resistant plants or to act as promoters,
 CC enhancers or termination signals.
 SQ Sequence 286 AA;
 Query Match 5.88; Score 136; DB 31; Length 286;
 Best Local Similarity 31.08; Pred. No. 7.38e-03;
 Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;
 Db 8 wmtinnpt-tlpv-mrd-eik-ykvqv-erg-qeg-trhvgvgvkmkrsslkqmrfg 60
 QY 17 WFTLNPNSEDERKKIRELPISLFDYFVIGEGNEEGRTPHLOGFANFVKQTEN 76
 Db 61 fpga-hlekrksgsearsymcke 83
 QY 77 LGARCHIEKAKGTDOQNKKEYCSKE 100
 RESULT 5
 ID W02157 standard; Protein; 611 AA.
 AC W02157;
 DT 14-JAN-1997 (first entry)

DE Periplasmic Beta-N-acetylglucosaminidase.
 KW Periplasmic chitinase; periplasmic Beta-N-acetylglucosaminidase;
 KW Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;
 KW catabolism.
 OS Vibrio furnissii.
 PN W09625424-AL.
 PD 22-AUG-1996. U02332.
 PF 13-FEB-1996; US-386727.
 PR 13-FEB-1995; US-386727.
 PA (UVO) UNIV JOHNS HOPKINS.
 PI Bassler B, Chittlaru E, Keyhani N, Roseman S, Rowe C;
 PI Yu C;
 DR WPI; 96-393335/39.
 DR N-PSDB; T36388.
 PT Chitin biosynthetic enzymes end I, exo I and exo II - are
 PT periplasmic chitinase(s), periplasmic beta-GlcNAcidase(s) and
 PT aryl beta-N-acetylglucosaminidase(s), respectively
 PS Claim 10; Page 73-75; 101pp; English.
 CC Periplasmic chitinase (W02156), periplasmic
 CC Beta-N-acetylglucosaminidase (W02157) and aryl
 CC Beta-N-acetylglucosaminidase (W02158) can be used to produce chitin
 CC oligosaccharides with the structure (GlcNAc)_n where n is 2 or
 CC higher, by contacting them with soluble chitin. The enzymes are
 CC encoded by the genes endI, exoI and exoII respectively. They are
 CC all genes involved in the catabolic pathway of chitin.
 SQ Sequence 611 AA;

Query Match 4.1%; Score 97; DB 19; Length 611;
 Best Local Similarity 25.3%; Pred. No. 1.22e+01;
 Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;
 Db 484 weeaqgdkvskdtviyswlsseqaln-carqgfdvllpqqgftldlaqdyapepgvd 542
 Qy 203 WD-GYGEKVVVDYFYGWLPWDDLLRLCDRLPLVTKGTVFPLARSILITSNQTPLE 261
 Db 543 w-agvt-p-lerayr 554
 Qy 262 WYSSTAVPAVEALYR 276

RESULT 6

ID W85599 standard; protein; 611 AA.
 AC W85599;
 DT 02-MAR-1999 (first entry)
 DT 08-MAR-1999 (revised)
 DE Hexosaminidase enzyme.
 KW Hexosaminidase; enzyme; laundry; cleaning agent; hydrolysis;
 KW anti-microbial; detergent; surfactant.
 OS unidentified.
 PN W09850512-AL.
 PD 12-NOV-1998.
 PF 05-MAY-1998; U09125.
 PR 19-AUG-1987; US-056132.
 PR 06-MAY-1997; US-045756.
 PA (PROC) PROCTER & GAMBLE CO.
 PI Convents AC, Moese RL, Wolff AM;
 DR WPI; 99-024116/02.
 DR N-PSDB; V83129.
 PT Laundry and cleaning compositions containing hexosaminidase - to
 PT provide antimicrobial activity and remove biofilm
 PS Claim 2; Page 38-39; 64pp; English.
 CC Novel hexosaminidase enzymes (W85599-605) can be used in
 CC combination in an aqueous laundry or cleaning product. The cleaning
 CC product is used especially used to launder fabrics and to clean
 CC dishes and tableware, particularly in an automatic dishwasher, but
 CC may also be used generally as hard surface cleaner. The cleaning
 CC product imparts antimicrobial activity and/or eliminates biofilm,
 CC the hexosaminidases having a minimum inhibitory concentration of
 CC less than about 0.125% but more preferably less than about 0.025%.
 CC (NB: entry was revised to change incorrect cross references in
 CC comments field).
 SQ Sequence 611 AA;

Query Match 4.1%; Score 97; DB 39; Length 611;
 Best Local Similarity 25.3%; Pred. No. 1.22e+01;
 Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;
 Db 484 weeaqgdkvskdtviyswlsseqaln-carqgfdvllpqqgftldlaqdyapepgvd 542
 Qy 203 WD-GYGEKVVVDYFYGWLPWDDLLRLCDRLPLVTKGTVFPLARSILITSNQTPLE 261
 Db 543 w-agvt-p-lerayr 554
 Qy 262 WYSSTAVPAVEALYR 276

RESULT 7

ID W62673 standard; protein; 165 AA.
 AC W62673;
 DT 09-NOV-1998 (first entry)
 DE Streptococcus pneumoniae polypeptide.
 KW Polypeptide; ORF; open reading frame; infection; bacterial;
 KW streptococcal; bacteremia; diagnosis; prophylaxis.
 OS Streptococcus pneumoniae.
 PN W09823631-AL.
 PD 04-JUN-1998.
 PF 24-NOV-1997; U21976.
 PR 27-NOV-1996; US-031879.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO,
 PI Reid RH, Jarfos EN;
 DR WPI; 98-322654/26.
 PT Streptococcus pneumoniae polynucleotides - useful for developing
 PT products for diagnosis, prevention and treatment of infections e.g.
 PT pneumonia, bacteremia, meningitis or endocarditis
 PS Claim 5; Page 32; 181pp; English.
 CC The sequence is that of a Streptococcal polypeptide.
 CC The polypeptide can potentially be used for the diagnosis and
 CC prevention of bacterial infections, especially SP infection.
 CC It may be used for the treatment of diseases such as otitis media,
 CC conjunctivitis, pneumonia, bacteremia, meningitis, sinusitis, pleural
 CC empyema, endocarditis or infection of the cerebrospinal fluid.
 SQ Sequence 165 AA;

Query Match 3.9%; Score 92; DB 34; Length 165;
 Best Local Similarity 18.9%; Pred. No. 2.94e+01;
 Matches 14; Conservative 27; Mismatches 29; Indels 4; Gaps 4;

Db 86 jpslkhqhsvvdtklktfttliqadnqrffidllifqiltkmngsmariarr-rlnwg 144
 Qy 115 QPSDLSTAVSTLLESGI-LVT-VAEQHPVTFVKNFRGLAELLKVGKMRDKNTNVHFI 172

Db 145 isrpacs-tkwans 157

Qy 173 VGPPGCGCKSKWAAN 186

RESULT 8

ID R47468 standard; protein; 820 AA.
 AC R47468;
 DT 03-AUG-1994 (first entry)
 DE Branching enzyme of rice starch.
 KW Rice; starch; transit peptide; pectin; cereal; aminopectin; seeds;
 KW reverse transcriptase; plaques.
 OS Cryza sativa.
 FH Key Location/Qualifiers
 FT peptide i.60 /label= transit_peptide
 FT J05317057-A.
 PN 03-DEC-1993.
 PD 30-MAR-1992; 102499.
 PR 20-SEP-1991; JP-48617.
 PA (MITS-) MITSUI GYOSAI SHOKUBUTU BIO KENKYUSHO KK.
 DR WPI; 94-011022/02.
 DR N-PSDB; 054f74.

PT Gene CDNA for rice starch branching enzyme for varied amino
 PT pectin in cereal - comprises structural gene specified by basic
 PT sequence introduced in rice plant for improved taste, for DNA
 PT fragment originated from rice genome contg. gene
 PS Claim 1; Page 5-7; 21pp; Japanese.
 CC The sequence shows a cDNA encoding a branching enzyme of rice
 CC starch. The enzyme can be used to modify amylopectin content of
 CC starch in cereal particles by introducing the basic sequence into
 CC a rice plant. This process can be used to improve the taste of the
 CC rice.
 SQ Sequence 820 AA;

Query Match 3.9%; Score 91; DB 9; Length 820;
 Best Local Similarity 40.6%; Pred. No. 3.50e+01;
 Matches 13; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 89 khdfnykrlyldqkcliekheggleefsky 120

QY 67 KQTFN-KVWYLGARCHIEKAKGT-DOONKEY 96

RESULT

ID R53228 standard; Protein; 820 AA.

AC R53228;

DT 06-MAR-1995 (first entry)

DE Rice starch branching enzyme.

KW Rice starch branching enzyme; oryza sativa; amylopectin; albumen;
 starch.

OS Oryza sativa.

PN J06098656-A.

PD 12-APR-1994.

PF 30-MAR-1992; 102500.

PR 30-MAR-1992; JP-102500.

PA (MITS-) MITSUI GYOSAI SHOKUBUTSU BIO KENKYUSHO.

DR WPI; 94-155835/19.

DR N-PSDB; Q62135.

PT Transgenic rice containing the rice starch branch family enzyme -
 used to increase the amylopectin content of albumen

PS Claim 3; Page 6-8; 24pp; Japanese.

CC The introduction of the rice starch branch-forming enzyme gene into
 a rice increase the activity of this enzyme in the plant, thereby

CC increasing the content of amylopectin in albumen starch and thus

CC enabling efficient mass production of various proteins.

SQ Sequence 820 AA;

Query Match 3.9%; Score 91; DB 11; Length 820;
 Best Local Similarity 40.6%; Pred. No. 3.50e+01;
 Matches 13; Conservative 8; Mismatches 9; Indels 2; Gaps 2;

Db 89 khdfnykrlyldqkcliekheggleefsky 120

QY 67 KQTFN-KVWYLGARCHIEKAKGT-DOONKEY 96

RESULT

ID R55273 standard; Protein; 1822 AA.

AC R55273;

DT 31-JAN-1995 (first entry)

DE Beta subunit of integrin cell surface receptor.

KW Integrin; alpha; beta; subunit; glycoprotein; heterodimer;

KW transmembrane; extracellular matrix; cell signalling; cytoskeleton;

KW behaviour; signal transduction; receptor.

OS Homo sapiens.

FH Key Location/Qualifiers

FT peptide 28..42 /note= "N-terminal peptide."

FT modified_site 491

FT /note= "Potential N-linked glycosylation site."

FT modified_site 617

FT /note= "Potential N-linked glycosylation site."

FT modified_site 695

FT /note= "Potential N-linked glycosylation site."

FT domain 711..733

FT modified_site 980 /note= "Putative transmembrane domain."
 FT /note= "Potential N-linked glycosylation site."
 FT modified_site 1593
 FT /note= "Potential N-linked glycosylation site."
 PN US5320942-A.
 PD 14-JUN-1994.
 PF 19-FEB-1987; 016552.
 PR 19-FEB-1987; US-016552.
 PR 04-JAN-1989; US-293384.
 PR 01-OCT-1990; US-591105.
 PA (KAJI/) KAJIJI S.
 PA (QUAK/) QUARANTA V.
 PI Kajiiji S.; Quaranta V.
 DR WPI; 94-191533/23.
 DR N-PSDB; Q65674.
 PT Diagnosing presence of abnormal epithelial tissue in vitro -
 PT utilises monoclonal antibodies to alpha6 beta4 cell surface protein
 PS Example 5; Figure 9; 34pp; English.
 CC Integrins are heterodimers comprised of alpha and beta subunits, that
 CC are non-covalently associated transmembrane glycoproteins. 11 alpha
 CC chains and 6 beta chains have been recognised in man. Each alpha
 CC subunit tends to associate with only one type of beta subunit but
 CC there are several exceptions. Integrins mediate (in part) the
 CC interaction of cells with the extracellular matrix, forming a link
 CC between the extracellular matrix and the cytoskeleton. They may
 CC transmit signals from the extracellular to the intracellular
 CC environment, affecting cell behaviour. This sequence is the beta4
 CC subunit of ah alpha6 beta4 integrin.
 SQ Sequence 1822 AA.

Query Match 3.9%; Score 92; DB 10; Length 1822;

Best Local Similarity 28.8%; Pred. No. 2.94e+01;

Matches 17; Conservative 19; Mismatches 19; Indels 4; Gaps 4;

Db 1013 irrvldgksgvyrtdgtagrdyvpvegellfpggeawkelqkllleqvdsll 1071

QY 73 VKWYL-GARCHIE-KAK-GTDQONKEYCSKEGNLLIECGAPRSQGRSGLS-TAVSTLL 127

RESULT

ID P60045 standard; Protein; 2150 AA.

AC P60045;

DT 30-JUL-1991 (first entry)

DE Sequence of viral proteins VP-1-VP4, P2A-P2C, P3A, VPg, protease

DE. polymerase.

KW Picornavirus; therapy.

OS Rhinovirus HRV2.

FH Key Location/Qualifiers

FT protein 1..69 /label= VP4

FT protein 70..330 /label= VP2

FT protein 331..567 /label= VP3

FT protein 568..855 /label= VP1

FT protein 856..881 /label= ?

FT protein 882..991 /label= ? P2-A

FT protein 992..1086 /label= P2-B

FT protein 1087..1409 /label= P2-C

FT protein 1410..1486 /label= P3-A

FT protein 1487..1507 /label= VPg

FT protein 1508..1690 /label= Protease

FT protein 1691..2150 /label= Polymerase

PN	EP-192175-A.
PD	27-AUG-1986.
PF	12-FEB-1986; 101746.
PR	15-FEB-1985; DE-505148.
PA	(BOEH) BOEHRINGER INGELHEIM.
PI	Stern T, Sommergruber W, Blaas D, Grundler P, Fraundorfer F,
PI	Duchlier M;
DR	WPI; 86-226788/35.
DR	N-PSDB; N60044.
PT	DNA coding for rhino:virus protein - and therapeutically
PT	effective polypeptide(s) encoded by this DNA
PS	Claim 11; Fig 4; 72pp; German.
CC	The polypeptides are useful for therapeutic treatment, e.g. for
CC	stimulating the immune system and for binding and/or blocking cell
CC	receptors for HRV2.
SQ	Sequence 2150 AA;
Query Match	3.9%; Score 93; DB 3; Length 2150; ..
Best Local Similarity	40.0%; Pred. No. 2.47e+01;
Matches	22; Conservative 12; Mismatches 14; Indels 7; Gaps 6;

```

Query Match      3.98; Score 93; DB 3; Length 2150; ...
Best Local Similarity 40.08; Pred. No. 2.47e+01;
Matches 22; Conservative 12; Mismatches 14; Indels 7; Gaps 6;

Db 1205 vaivhpgpgagkxi-ttnflakmindsdiyalpdpkyfdgdyqgsvvimmddi 1258
QY 169 VHFIV-gpccgksgkwaane-AN-P-ET-TYKPKNKWDCYHGEKVYVVDFF 217

```

RESULT	ID	Accession	Description	Key	Location/Qualifiers
12	P81048	standard; protein; 2150 AA.			
AC	P81048;				
DE	22-MAR-1991	(first entry)			
DE	Sequence encoded by Rhinovirus 2 (HRV2)	genome cDNA			
KW	Passive immunity; diagnosis; therapy; ss.				
OS	Rhinovirus.				
FH	key				
FT	protein	1..69			/label=VP4
FT	protein	70..330			/label=VP2
FT	protein	331..567			/label=VP3
FT	protein	568..856			/label=VP1
FT	protein	857..882			/label=?
FT	protein	883..992			/note="protein named '?'"
FT	protein	993..1087			/label=P2-A
FT	protein	1088..1409			/label=P2-B
FT	protein	1410..1486			/label=P2-C
FT	protein	1487..1507			/label=P3-A
FT	protein	1508..1690			/label=VPg
FT	protein	1691..2150			/label=PROTEASE
FT	protein	1691..2150			/label=POLYMERASE

EP-	261403-A.
PN	
PD	30-MAR-1988.
PD	
PF	20-AUG-1987; 112104.
PF	
PR	17-JAN-1987; DE-701301.
PR	(BOEH) Boehringer Ingelheim.
PI	Duechler M, Skern T, Sommergruber W, Neubauer C, Grundler P, Blaas,
PI	Kuchler E, Frasei L, Zorn M;
PI	WPI: 88-085735/13.
DR	N-PSDB; N81393.
DR	
PT	New DNA corresponding to viral RNA of rhino-virus HRV89 - useful for
PT	prodn. of polypeptide(s) for stimulating immune system against HRV
PT	89.
PT	
PS	Example: Fig 14; 66pp; German..
PS	
CC	DNA molecules corresponding to all or part of the RNA of rhinovirus
CC	

CC strain HRV89 (Fig 4, N91390) is claimed, esp. the portion encoding
CC the v1- α l proteins VP1-VP4, P2A-P2C, P3A-P3C. Also claimed are the
CC polypeptides encoded by any of these DNA molecules. The polypeptides
CC are used for stimulating a protective immune response and for
CC blocking cellular receptors. Ab are useful for assay and for
CC purific. of the corresp. antigen, and can also be used for the
CC therapeutic and diagnostic applications.
CC sequence '2150 AA;
SQ

```

Query Match:          3.9%; Score 93; DB 2; Length 2150;
Best Local Similarity 40.0%; Pred. No. 2.47e+01;
Matches 22; Conservative 14; Mismatches 14; Indels 7; Gaps 6;

Db 1205 vaivihgppgagkssi-ttnflakmitndsiyslppdkyfgdgqsvimddi 1258
Qv 169 VHFIV-PPGCGKSKWAAN-AN--P-ET-TWYKPPKNKWDPDYGHEKVVVDDF 217

```

RESULT	13	
ID	R05127	standard; protein; 2150 AA.
AC	R05127;	
DT	17-JUL-1990	(first entry)
DE	Complete human rhinovirus 2.	
KW	Human rhinovirus 2.	
FH	Key	Location/Qualifiers
FT	peptide	1..69
FT	peptide	/label-VP4
FT	peptide	70..330
FT	peptide	/label-VP2
FT	peptide	331..567
FT	peptide	/label-VP3
FT	peptide	568..852
FT	peptide	/label-VP1
FT	peptide	853..967
FT	peptide	/label-P2-A
FT	peptide	993..1087
FT	peptide	/label-P2-B
FT	peptide	1088..1486
FT	peptide	/label-P2-C
FT	peptide	1487..1507
FT	peptide	/label-VPg
FT	protein	1508..1690
FT	protein	/label-Protease
FT	protein	1691..2150
FT	protein	/label-Polymerase

PN	W99001061-A.
PD	08-FEB-1990.
PF	22-JUL-1989; 0000861.
PR	24-JUN-1989; DE-920753.
PR	24-JUN-1989; DE-825189.
PI	(BOEH) Boehringer Ingelheim.
PI	Duelcher M, Skern T, Blaas D, Berger B, Sommergruber W, Keuchler E;
DR	WPI: 90-067175/09.
DR	N-PSDB: 003418.
PT	New plasmid containing complete cDNA of human rhinovirus 2 under control
PT	of RNA polymerase promoter, transcribable in vitro to infectious RNA.
PS	Disclosure; Fig.3A-3E; 33pp; German.
CC	The HRV2 cDNA transcribed to infectious RNA allows study of
CC	viral behaviour. Effects of deletions, insertions or exchanges of amino
CC	acids can be examined directly and quantitatively. Viral polypeptides
CC	expressed by these plasmids can be used therapeutically, eg. to stimulate
CC	the immune system.
CC	Sequence 2150 AA.
SQ	Sequence 2150 AA.

```

Query Match      , 3.9%;      Score 93;  DB 2;  Length 2150;
Best Local Similarity 40.0%;
Matches 22;  Conservative 12;  Mismatches 14;  Indels 7;  Gaps 6;

Db 1205.vai.vihgppgagkai-tcnflakitndslyslppdpkyfdgqgsvvmdidi 1258
Qv 169 yHFIV-gpppgckrskwaanF-AN--P-ET-YWKKPPKKNWDDYGEKVVVDDF 217
      :::::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::

```


RESULT 14
ID P60066 standard; Protein; 2227 AA.

AC P60066;
DT 26-JUN-1991 (first entry)
DE Sequence of viral I434 polypeptide encoded by the complete
DE nucleotide sequence of the HAV genome.
KW Diagnosis; vaccine; passive immunotherapy.
OS Hepatitis A virus.

FH Key Location/Qualifiers
FT region 1..245

FT FT /label= P1.1A

FT FT 246..491

FT FT /label= 1B

FT FT 492..836

FT FT /label= 1C

FT FT 837..980

FT FT /label= P2.2A

FT FT 981..1076

FT FT /label= 2B

FT FT 1077..1422

FT FT /label= 2C

FT FT 1423..1484

FT FT /label= P3.3A

FT FT 1485..1507

FT FT /label= 3B

FT FT 1508..1678

FT FT /label= 3C

FT FT 1679..2227

FT FT /label= 3D

PN EP-199480-A.

PD 29-OCT-1986.

PF 03-APR-1986; 302465.

PR 03-APR-1985; US-719329.

PA (CHIR-) CHIRON CORP.

PI Dina D, Potter SJ, Vannest GA, Caput D;

DR WPI: 86-286213/44.

DR N-PSDB; N60080.

DR Hepatitis A virus nucleotide sequence and polypeptide - and use

PT in prodn. of vaccines and diagnostic probes

PS Claim 5; Fig 1; 18pp; English.

CC N60080 and oligonucleotide fragments are useful in detection of

CC hepatitis A virus; transformed hosts may be used for expression of

CC polypeptides and fragments useful in vaccines without risk of

CC infection by the virus or in prodn. of particles which are capable

CC of inducing immunocompetent B cells for passive immunotherapy. Pref.

CC epitope is derived from AAs 445-657 or 792-848 of the HAV

CC polypeptide sequence (P60066).

SQ Sequence 2227 AA;

Query Match 3.9%; Score 92; DB 3; Length 2227;

Best Local Similarity 46.7%; Pred. No. 2.94e+01;

Matches 14; Conservative 5; Mismatches 10; Indels 1; Gaps 1;

Db 1254 epeknlytkpvasydydgysgqlvcilddi 1283

QY 189 NPETT-YWKPPKKNWDCYHGEKVVIDDF 217

RESULT 15

ID W34074 standard; Protein; 2227 AA.

AC W34074;

DT 27-APR-1998 (first entry)

DE Hepatitis A virus HM-175 protein sequence.

DE HAV; attenuation; 2C protein; 2C chimeric hepatitis A virus;

KW infection; vaccine.

OS Hepatitis A virus HM-175.

FH Key Location/Qualifiers

FT Protein 1..23

FT FT /label= VP4

FT FT 24..245

FT FT /label= VP2

FT FT 246..491

FT FT /label= VP3

FT Protein 492..791

FT FT /label= VP1

FT Protein 792..980

FT FT /label= 2A

FT Protein 981..1087

FT FT /label= 2B

FT Protein 1088..1422

FT FT /label= 2C

FT Protein 1423..1496

FT FT /label= 3A

FT Protein 1497..1519

FT FT /label= 3B

FT Protein 1520..1738

FT FT /label= 3C

FT Protein 1739..2227

FT FT /label= 3D

PN W09740166-A2.

PD 30-OCT-1997.

PF 18-APR-1997; U06506.

PR 19-APR-1996; US-015642.

PA (USSH) US SEC DEPT HEALTH...

PA (USSH) US DEPT HEALTH & HUMAN SERVICES.

PI Emerson SU, Purcell RH, Raychaudhuri G;

DR WPI: 97-535850/49.

DR N-PSDB; T93023.

PT Human attenuated HAV genome containing simian HAV 2C gene - useful

PT as vaccines against HAV infection

PS Disclosure; Fig 13A-D; 66pp; English.

CC This protein sequence is encoded by the human hepatitis A virus

CC (HAV) HM-175 wild-type genome (see T93023). Attenuated strain

CC HAV/7 is obtained by passage of HM-175 in African Green Monkey

CC kidney cells. A claimed DNA construct (1) comprises a genome of

CC HAV, where the genome is a human attenuated HAV genome in which a

CC region of the 2C gene has been replaced by a corresponding region

CC from a 2C gene of a simian AGM-27 HAV genome (see T93024). The

CC region of the 2C gene from AGM-27 contained in the construct

CC preferably encodes amino acids 120-328 of the 2C protein, amino

CC acids 1-121 or amino acids 1-328. Also claimed are: (1) an RNA

CC transcript of (1); (2) a cell transfected with (1) or the RNA

CC transcript of (1); (3) a HAV genome as above; (4) antibodies to the

CC HAV of (3); and (5) a host cell containing the HAV of (3). (1) or

CC its RNA transcript, can be used as a vaccine for preventing HAV in

CC a mammal. (1) or the RNA transcript can also be used to stimulate

CC the production of protective antibodies in the mammal.

SQ Sequence 2227 AA;

Query Match 3.9%; Score 92; DB 27; Length 2227;

Best Local Similarity 46.7%; Pred. No. 2.94e+01;

Matches 14; Conservative 5; Mismatches 10; Indels 1; Gaps 1;

Db 1254 epeknlytkpvasydydgysgqlvcilddi 1283

QY 189 NPETT-YWKPPKKNWDCYHGEKVVIDDF 217

Search completed: Wed Dec 22 10:11:49 1999

Job time : 26 secs.

 W P S R L H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

Mpsrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:14:00 1999; Maspar time 6.62 seconds
 Tabular output not generated. 565.408 Million cell updates/sec

Title: >US-09-209-961-3
 Description: (1-314) from US09209961.pep
 Perfect Score: 2358
 Sequence: 1 MPKNGRSGPQPKRWFT.....GGQFVTLSPPCPEPPEYNY 314

Scoring table: PAM 150
 Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: a-issued
 1:5A_COMB 2:5B_COMB 3:PCT9_COMB 4:backfiles1
 Statistics: Mean 32.212; Variance 134.292; scale 0.240

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	242	10.3	290	2	US-08-202-Sequence 26, Applicati	2.13e-12
2	197	8.4	286	2	US-08-418-Sequence 16, Applicati	2.67e-08
3	193	8.2	285	2	US-08-418-Sequence 17, Applicati	6.10e-08
4	139	5.9	286	2	US-08-418-Sequence 18, Applicati	3.02e-03
5	136	5.8	286	2	US-08-202-Sequence 25, Applicati	5.38e-03
6	97	4.1	611	2	US-08-386-Sequence 4, Applicatio	7.06e+00
7	89	3.8	445	2	US-08-630-Sequence 4, Applicatio	2.78e+01
8	89	3.8	445	2	US-08-838-Sequence 4, Applicatio	2.78e+01
9	89	3.8	456	3	PCT-US95-1-Sequence 2, Applicatio	2.78e+01
10	89	3.8	456	1	US-08-349-Sequence 2, Applicatio	2.78e+01
11	90	3.8	1463	1	US-08-157-Sequence 3, Applicatio	2.35e+01
12	88	3.7	445	2	US-08-630-Sequence 6, Applicatio	3.29e+01
13	88	3.7	445	2	US-08-838-Sequence 6, Applicatio	3.29e+01
14	88	3.7	455	1	US-08-349-Sequence 4, Applicatio	3.29e+01
15	88	3.7	456	3	PCT-US95-1-Sequence 4, Applicatio	3.29e+01
16	88	3.7	533	1	US-08-488-Sequence 6, Applicatio	3.29e+01
17	86	3.6	346	2	US-08-744-Sequence 2, Applicatio	4.60e+01
18	86	3.6	428	4	5179023-4 Patent No. 5179023	4.60e+01
19	86	3.6	429	1	US-08-261-Sequence 7, Applicatio	4.60e+01
20	86	3.6	429	1	US-07-983-Sequence 2, Applicatio	4.60e+01
21	86	3.6	429	1	US-07-602-Sequence 3, Applicatio	4.60e+01
22	86	3.6	429	1	US-07-602-Sequence 2, Applicatio	4.60e+01
23	86	3.6	429	1	US-08-261-Sequence 9, Applicatio	4.60e+01

24	86	3.6	429	1	US-08-261-Sequence 3, Applicatio	4.60e+01
25	83	3.5	627	2	US-08-700-Sequence 6, Applicatio	7.55e+01
26	83	3.5	627	2	US-08-466-Sequence 6, Applicatio	7.55e+01
27	83	3.5	1724	1	US-08-325-Sequence 3, Applicatio	7.55e+01
28	81	3.4	310	3	PCT-US95-0-Sequence 42, Applicati	1.05e+02
29	81	3.4	310	3	PCT-US95-0-Sequence 42, Applicati	1.05e+02
30	81	3.4	310	2	US-08-337-Sequence 42, Applicati	1.05e+02
31	81	3.4	310	2	US-08-433-Sequence 42, Applicati	1.05e+02
32	80	3.4	325	1	US-08-233-Sequence 55, Applicati	1.23e+02
33	81	3.4	346	1	US-08-458-Sequence 2, Applicatio	1.05e+02
34	91	3.4	346	1	US-08-213-Sequence 2, Applicatio	1.05e+02
35	91	3.4	346	1	US-08-747-Sequence 2, Applicatio	1.05e+02
36	81	3.4	346	1	US-08-460-Sequence 2, Applicatio	1.05e+02
37	91	3.4	346	1	US-08-299-Sequence 6, Applicatio	1.05e+02
38	91	3.4	365	2	US-08-089-Sequence 5, Applicatio	1.05e+02
39	91	3.4	365	2	US-08-421-Sequence 5, Applicatio	1.05e+02
40	91	3.4	365	1	US-08-421-Sequence 5, Applicatio	1.05e+02
41	91	3.4	365	2	US-08-851-Sequence 4, Applicatio	1.05e+02
42	80	3.4	480	2	US-08-833-Sequence 49, Applicatio	1.23e+02
43	80	3.4	480	1	US-08-306-Sequence 49, Applicati	1.23e+02
44	80	3.4	480	1	US-08-476-Sequence 49, Applicati	1.23e+02
45	80	3.4	507	1	US-08-457-Sequence 23, Applicati	1.23e+02

ALIGNMENTS

RESULT 1
 ID US-08-202-186-26, STANDARD; PRT; 290 AA.

XX XXXXXX

Sequence 26, Application US/08202186

Sequence 26, Application US/08202186

Patent No. 5756708

GENERAL INFORMATION:

APPLICANT: KARAN, Mirko

APPLICANT: BURNS, Thomas M.

APPLICANT: DALE, James L.

TITLE OF INVENTION: HARDING, Robert M.

NUMBER OF SEQUENCES: 60

CORRESPONDENCE ADDRESS:

ADDRESSEE: Foley & Lardner

STREET: 3000 K Street, N.W.

CITY: Washington, D.C.

COUNTRY: USA

ZIP: 20007-5109

COMPUTER: READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/202,186

FILING DATE: 24-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: JEFFERY, Donald D.

REGISTRATION NUMBER: 19,980

REFERENCE/DOCKET NUMBER: 71611/102 FIVE

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 672 5300

TELEFAX: 202 672 5399

TELEX: 904136

INFORMATION FOR SEQ ID NO: 26:

SEQUENCE CHARACTERISTICS:

LENGTH: 290 amino acids

TYPE: amino acid

TOPOLOGY: linear

SEQUENCE 290 AA; 33497 MW; 430051 CN;

Query Match 10.3%; Score 242; DB 2; Length 290;
Best Local Similarity 38.2%; Pred. No. 2.13e-12;
Matches 42; Conservative 26; Mismatches 36; Indels 6; Gaps 6;

DB 6 RWCFTLNTVETEEAANVVRISLNLV-YAIVGDEVAPSTGQRHLQGFTHLKTGRRLQG 64
QY 15 KRWFTLNPSEDERKKI-REL-PISLFYFVGEENEGRTPHLOGFANFVKKQTFNK 72
DB 65 LKTVLGNDRHLPTGRSGDEQNDYCSKE-RVLLHGVPTPRPGVKRPLA 113
QY 73 VKWYLGAR-CHIEKAKGTDOONKEYCSKEGNLLIEGAPRSQG-QRSDL 120

RESULT 2
ID US-08-418-071-16 STANDARD; PRT; 286 AA.

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

Sequence 16, Application US/08418071

Patent No. 5846705

GENERAL INFORMATION:

APPLICANT: Wu, Rey-Yuh

APPLICANT: You, Li-Ru

TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA

TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DET

TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Neave

STREET: 1251 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States

ZIP: 10020

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Word Perfect 5.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/418,071

FILING DATE: 06-APR-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haley Jr., James F.

REGISTRATION NUMBER: 27,794

REFERENCE/DOCKET NUMBER: DCB-1

TELEPHONE: (212) 596-9000

TELEFAX: (212) 596-9090

Query Match 8.4%; Score 197; DB 2; Length 286;
Best Local Similarity 36.5%; Pred. No. 2.67e-08;
Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

DB 4 PSLKWCFTLNTVETEEAANVVRISLNLV-YAIVGDEVAPSTGQRHLQGFTHLKTGRRLQG 63
QY 13 PKRWFTLNPSEDERKKI-REL-PISLFYFVGEENEGRTPHLOGFANFVKKQTFNK 71

RESULT 2
ID US-08-418-071-16 STANDARD; PRT; 286 AA.

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

CC STATE: D.C.
CC COUNTRY: USA
CC ZIP: 20005
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/386,727
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: HOBBS, ANN S.
CC REGISTRATION NUMBER: 36,830
CC REFERENCE/DOCKET NUMBER: 4130/206916
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-822-0944
CC TELEX: 6714627 CUSH
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 611 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 611 AA; 69501 MW; 1856094 CN;

Query Match 4.1%; Score 97; DB 2; Length 611;
Best Local Similarity 25.38; Pred. No. 7.06e+00;
Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;

Db 484 WEQAQHGKVKDTIYISWLSQAALN-CARQGFDTILQPGQFTYLDIAQDYAPEEPGVD 542
QY 203 WD-GYHGEKVVIDDFYGLPWDLRLCDRYPLTVTKGTVPFLARSILITSNQIPPLE 261
Db 543 W-AGVT-P-LERAYR 554
QY 262 WYSTAVPAVEALYR 276

RESULT 7
ID US-08-630-118A-4 STANDARD; PRT; 445 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 4, Application US/08630118A
Sequence 4, Application US/08630118A
Patent No. 5919901
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive, 32nd Floor
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 4:

CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/630,118A
CC FILING DATE: April 8, 1996
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Greenfield Ph.D., Michael S.
CC REGISTRATION NUMBER: 37,147
CC REFERENCE/DOCKET NUMBER: 96,149/WH 405
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312)715-1000
CC TELEFAX: (312)715-1234
CC INFORMATION FOR SEQ ID NO: 4:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 445 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 445 AA; 50408 MW; 1091793 CN;

Query Match 3.8%; Score 89; DB 2; Length 445;
Best Local Similarity 39.4%; Pred. No. 2.78e+01;
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 49 TFSVLLGFMGNLLILMAVKKRNOKTTVNFILG 81
QY 142 TFSVLLGFMGNLLILMAVKKRNOKTTVNFILG 174

RESULT 8
ID US-08-838-399-4 STANDARD; PRT; 445 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 4, Application US/08838399
Sequence 4, Application US/08838399
Patent No. 5965392
GENERAL INFORMATION:
APPLICANT: Hu Ph.D., Yinghe
APPLICANT: McCaleb Ph.D., Michael L.
APPLICANT: Bloomquist Ph.D., Brian T.
APPLICANT: Flores-Riveros Ph.D., Jaime R.
APPLICANT: Cornfield Ph.D., Linda J.
TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
TITLE OF INVENTION: Sequences
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
STREET: 300 South Wacker Drive
CITY: Chicago
STATE: IL
COUNTRY: USA
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/838,399
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Greenfield Ph.D., Michael S.
REGISTRATION NUMBER: 37,147
REFERENCE/DOCKET NUMBER: 96,149/WH 405
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312)715-1000
TELEFAX: (312)715-1234
INFORMATION FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 445 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 445 AA; 50408 MW; 1091793 CN;

Query Match 3.8%; Score 89; DB 2; Length 445;
Best Local Similarity 39.4%; Pred. No. 2.78e+01;
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 49 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 81
||| : : : : : ||| : : : : :
QY 142 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 174

RESULT 9
ID PCT-US95-15646-2 STANDARD; PRT; 456 AA.
XX xxxxxx

CC SEQUENCE 2, Application PC/TUS9515646
CC GENERAL INFORMATION:
CC APPLICANT: Synaptic Pharmaceutical Corporation
CC TITLE OF INVENTION: METHODS OF MODIFYING FEEDING BEHAVIOR, COMPOUNDS
CC TITLE OF INVENTION: USEFUL IN SUCH METHODS, AND DNA ENCODING A HYPOTHALAMIC
CC TITLE OF INVENTION: Y/PEPTIDE YY RECEPTOR (Y5) AND USES THEREOF
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham LLP
CC STREET: 1185 Avenue of the Americas
CC CITY: New York
CC STATE: New York
CC COUNTRY: United States of America
CC ZIP: 10036

CC COMPUTER READABLE FORM:
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/15646
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: White, John P.
CC REGISTRATION NUMBER: 28,678
CC REFERENCE/DOCKET NUMBER: 1795/46166-A-PCT
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 278-0400
CC TELEFAX: (212) 391-0525
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 456 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 456 AA; 51715 MW; 1146620 CN;

Query Match 3.8%; Score 89; DB 3; Length 456;
Best Local Similarity 39.4%; Pred. No. 2.78e+01;
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 60 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 92
||| : : : : : ||| : : : : :
QY 142 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 174

RESULT 10

US-08-349-025-2 STANDARD; PRT; 456 AA.
XX xxxxxx

Sequence 2, Application US/08349025

Sequence 3, Application US/08349025
Patent No. 5602024

GENERAL INFORMATION:

APPLICANT: Gerald, Christophe P.G.

APPLICANT: Walker, Mary W.

APPLICANT: Branchek, Theresa

APPLICANT: Weinshank, Richard L.

TITLE OF INVENTION: DNA ENCODING A HYPOTHALAMIC ATYPICAL

TITLE OF INVENTION: NEUROPEPTIDE Y/PEPTIDE YY RECEPTOR (Y5) AND USES THEREOF

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooper & Dunham

STREET: 1185 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/349,025

FILING DATE:

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: White, John P.

REGISTRATION NUMBER: 28,678

REFERENCE/DOCKET NUMBER: 1795/46166

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 278-0400

TELEFAX: (212) 391-0525

TELEX: 422523 COOP UI

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 456 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 456 AA; 51715 MW; 1146620 CN;

Query Match 3.8%; Score 89; DB 1; Length 456;
Best Local Similarity 39.4%; Pred. No. 2.78e+01;
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 60 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 92
||| : : : : : ||| : : : : :
QY 142 TFSVLLGFMGNLLILMAVMKKRQKTTVNFLLIG 174

RESULT 11
US-08-157-005-3 STANDARD; PRT; 1463 AA.

XX xxxxxx

Sequence 3, Application US/08157005

Sequence 3, Application US/08157005

Patent No. 5620891

GENERAL INFORMATION:

APPLICANT: Wensvoort, Gert

APPLICANT: Terpstra, Catharinus

CC APPLICANT: Pol, Johannes M
CC APPLICANT: Moorman, Robertus J
CC APPLICANT: Meulenbergh, Johanna J
CC TITLE OF INVENTION: CAUSITIVE AGENT OF THE MYSTERY SWINE DISEASE,
CC TITLE OF INVENTION: VACCINE COMPOSITIONS AND DIAGNOSTIC KITS
CC NUMBER OF SEQUENCES: 9
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Cooper & Dunham
CC STREET: 30 Rockefeller Plaza
CC CITY: New York
CC STATE: New York
CC COUNTRY: USA
CC ZIP: 10112
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.24
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/157,005
CC FILING DATE: 26-NOV-1993
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 91201398.4
CC FILING DATE: 06-JUN-1991
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 92200781.0
CC FILING DATE: 18-MAR-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/NL92/00096
CC FILING DATE: 05-JUN-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Moran, Thomas F
CC REGISTRATION NUMBER: 16,579
CC REFERENCE/DOCKET NUMBER: 44819
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 977-9550
CC TELEX: 422523 COOP UI
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1463 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1463 AA; 161876 MW; 11169848 CN;

Query Match 3.8%; Score 90; DB 1; Length 1463;
Best Local Similarity 44.0%; Pred. No. 2.35e+01;
Matches 11; Conservative 4; Mismatches 9; Indels 1; Gaps 1;

Db 791 FIVGPGSGKTTWLSQVQDDVIY 815
QY 171 FIVGPGCGKSKWA-ANFANPEITY 194

RESULT 12
ID US-08-630-118A-6 STANDARD; PRT; 445 AA.
XX XXXXXX
DT
XX
DE Sequence 6, Application US/08630118A
XX Sequence 6, Application US/08630118A
CC Patent No. 5919901
CC GENERAL INFORMATION:
CC APPLICANT: Hu Ph.D., Yinghe
CC APPLICANT: McCaleb Ph.D., Michael L.
CC APPLICANT: Bloomquist Ph.D., Brian T.
CC APPLICANT: Flores-Riveros Ph.D., Jaime R.
CC TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
CC STREET: 300 South Wacker Drive
CC CITY: Chicago
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

CC TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
CC TITLE OF INVENTION: Sequences
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
CC STREET: 300 South Wacker Drive, 32nd Floor
CC CITY: Chicago
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/630,118A
CC FILING DATE: April 8, 1996
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Greenfield Ph.D., Michael S.
CC REGISTRATION NUMBER: 37,142
CC REFERENCE/DOCKET NUMBER: 96,149/WH 405
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (312)715-1000
CC INFORMATION FOR SEQ ID NO: 6:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 445 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 445 AA; 50726 MW; 1109569 CN;

Query Match 3.7%; Score 88; DB 2; Length 445;
Best Local Similarity 39.4%; Pred. No. 3.29e+01;
Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 49 TSVSLGLFMGNLILMALMKRNQKTTVNFLLIG 81
QY 142 TFFVNFRLAEELLVSGMKQKRDWKTNVHFIIVG 174

RESULT 13
ID US-08-838-399-6 STANDARD; PRT; 445 AA.
XX XXXXXX
DT
XX
DE Sequence 6, Application US/08838399
XX Sequence 6, Application US/08838399
CC Patent No. 5965392
CC GENERAL INFORMATION:
CC APPLICANT: Hu Ph.D., Yinghe
CC APPLICANT: McCaleb Ph.D., Michael L.
CC APPLICANT: Bloomquist Ph.D., Brian T.
CC APPLICANT: Flores-Riveros Ph.D., Jaime R.
CC TITLE OF INVENTION: Neuropeptide Y Receptor and Nucleic Acid
CC NUMBER OF SEQUENCES: 8
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
CC STREET: 300 South Wacker Drive
CC CITY: Chicago
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60606
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/838,399

CC FILING DATE:

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Greenfield Ph.D., Michael S.

CC REGISTRATION NUMBER: 37,147

CC REFERENCE/DOCKET NUMBER: 96,149/WH 405

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (312)715-1000

CC TELEFAX: (312)715-1234

CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 445 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 445 AA; 50726 MW; 1109569 CN;

Query Match 3.7%; Score 88; DB 2; Length 445;

Best Local Similarity 39.4%; Pred. No. 3.29e+01;

Matches 13; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Db 49 TFSVLLGFMGNLLILMALMKRKNQKTTVNFLLIG 81

||| : : : : : ||| : : : : :

Qy 142 TFSVLLGFMGNLLILMALMKRKNQKTTVNFLLIG 174

RESULT 14

ID US-08-349-025-4 STANDARD; PRT; 455 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC INFORMATION FOR SEQ ID NO: 4:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 455 amino acids

CC TYPE: amino acid

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC SEQUENCE 455 AA; 51990 MW; 1160219 CN;

Query Match

Best Local Similarity

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Matches

Tue Jan 4 11:52:26 2000

US-09-209-961-3.rai

Page 8

Search completed: Wed Dec 22 10:14:10 1999
Job time : 10 secs.

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	242	10.3	290	A46353	ORF1 protein - coconu	2.95e-26
2	136	5.8	286	JQ1367	hypothetical 33.6k pr	1.09e-06
3	130	5.5	3085	T00327	polythetinal - infecti	1.04e-05
4	119	5.0	148	S04807	hypothetical protein,	5.62e-04
5	116	4.9	295	D24452	C1 protein - tobacco	1.61e-03
6	102	4.3	290	C71416	probable ABC transpor	1.78e-01
7	99	4.2	1864	J01657	genome polyprotein B	4.61e-01
8	100	4.2	2157	GNNY1B	genome polyprotein -	3.36e-01
9	97	4.1	244	JE0170	dnaj heat shock prote	8.61e-01
10	96	4.1	273	S04125	chlorophyll a/b-bind	1.17e+00
11	96	4.1	417	A64380	phosphoglycerate kina	1.17e+00
12	96	4.1	1777	T00490	nonstructural protein	1.17e+00
13	95	4.0	94	H70654	hypothetical protein	1.59e+00
14	95	4.0	286	S17696	chlorophyll a/b-bind	1.59e+00
15	95	4.0	415	S33627	ribulose-bisphosphate	1.59e+00
16	94	4.0	503	ALRPP	alpha-amylase [EC 3.2	2.16e+00
17	91	3.9	136	T00044	hypothetical 15.8k pr	5.28e+00
18	91	3.9	159	J83299	sigma-E factor regula	5.28e+00
19	91	3.9	211	S34274	ANG12 protein precurs	5.28e+00
20	91	3.9	243	S24559	clathrin light chain	5.28e+00
21	91	3.9	248	A31775	clathrin light chain	5.28e+00
22	91	3.9	286	LRR1A1	clathrin light chain	5.28e+00
23	92	3.9	313	H70777	probable cobD - Mycob	3.93e+00

∴ virus genome contains a putative replicase gene

```

#accession J01960
#molecule_type mRNA
#residues 1-286 #label HAR
KEYWORDS
FEATURE P-loop
180-188
SUMMARY
#region nucleotide-binding motif A (P-loop)
#length 286 #molecular-weight 33603 #checksum 3909

Query Match 5.8%; Score 136; DB 2; Length 286;
Best Local Similarity 31.0%; Pred. No. 1.09e-06;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMTINPT-TLPV-MED-EIK-YMYQV-ERG-QEG-TRHVGQVEMKRSSLKQMRGF 60
QY 17 WFTLNPPSEDERKKIRLPISFDYFIVGEEGNEGRTPHLOGFANFVKQTGNKVKY 76
Db 61 FPGALHLEKRRKGOEARSYCMKE 83
QY 77 LGARCHIEKAKGTDQNKVCSKE 100

RESULT 3
ENTRY T00327 #type complete
TITLE polyprotein - infectious flacherie virus
ORGANISM #formal_name infectious flacherie virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 29-Jan-1999
ACCESSIONS T00327
REFERENCE Z14139
#authors Isawa, H.; Asano, S.; Sahara, K.; Iizuka, T.; Bando, H.
#journal Arch. Virol. (1998) 143:127-143
#title Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses.

#accession T00327
#status preliminary
#residues 1-3085 #label ISA
#cross-references EMBL:AB000906; NID:d1185277; PID:d1026299
SUMMARY #length 3085 #molecular-weight 345801 #checksum 4939

Query Match 5.5%; Score 130; DB 3; Length 3085;
Best Local Similarity 35.3%; Pred. No. 1.04e-05;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PGKFEHPNPTVTRNPTSQWNGYNGQPIVLYDDI 1444
QY 184 AANFANPETTYWRPKKWDGYHGERVVIDDF 217

RESULT 4
ENTRY S04807 #type complete
TITLE hypothetical protein, 17.2K - maize streak virus (South African isolate)
ORGANISM #formal_name maize streak virus
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Sep-1997
ACCESSIONS S04807
REFERENCE S04804
#authors Lazarowitz, S.G.
#journal Nucleic Acids Res. (1988) 16:229-249
#title Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.
#cross-references MUID:88124198
#accession S04807
#status translation not shown
#molecule_type DNA
#residues 1-148 #label LAZ
#cross-references EMBL:Y00514; NID:g59365; PID:g59369
SUMMARY #length 148 #molecular-weight 17214 #checksum 3101

Query Match 5.0%; Score 119; DB 2; Length 148;
Best Local Similarity 26.7%; Pred. No. 5.62e-04;

#accession J01960
#molecule_type mRNA
#residues 1-286 #label HAR
KEYWORDS
FEATURE P-loop
180-188
SUMMARY
#region nucleotide-binding motif A (P-loop)
#length 286 #molecular-weight 33603 #checksum 3909

Query Match 5.8%; Score 136; DB 2; Length 286;
Best Local Similarity 31.0%; Pred. No. 1.09e-06;
Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMTINPT-TLPV-MED-EIK-YMYQV-ERG-QEG-TRHVGQVEMKRSSLKQMRGF 60
QY 17 WFTLNPPSEDERKKIRLPISFDYFIVGEEGNEGRTPHLOGFANFVKQTGNKVKY 76
Db 61 FPGALHLEKRRKGOEARSYCMKE 83
QY 77 LGARCHIEKAKGTDQNKVCSKE 100

RESULT 3
ENTRY T00327 #type complete
TITLE polyprotein - infectious flacherie virus
ORGANISM #formal_name infectious flacherie virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change 29-Jan-1999
ACCESSIONS T00327
REFERENCE Z14139
#authors Isawa, H.; Asano, S.; Sahara, K.; Iizuka, T.; Bando, H.
#journal Arch. Virol. (1998) 143:127-143
#title Analysis of genetic information of an insect picorna-like virus, infectious flacherie virus of silkworm: evidence for evolutionary relationships among insect, mammalian and plant picorna(-like) viruses.

#accession T00327
#status preliminary
#residues 1-3085 #label ISA
#cross-references EMBL:AB000906; NID:d1185277; PID:d1026299
SUMMARY #length 3085 #molecular-weight 345801 #checksum 4939

Query Match 5.5%; Score 130; DB 3; Length 3085;
Best Local Similarity 35.3%; Pred. No. 1.04e-05;
Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PGKFEHPNPTVTRNPTSQWNGYNGQPIVLYDDI 1444
QY 184 AANFANPETTYWRPKKWDGYHGERVVIDDF 217

RESULT 4
ENTRY S04807 #type complete
TITLE hypothetical protein, 17.2K - maize streak virus (South African isolate)
ORGANISM #formal_name maize streak virus
DATE 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 09-Sep-1997
ACCESSIONS S04807
REFERENCE S04804
#authors Lazarowitz, S.G.
#journal Nucleic Acids Res. (1988) 16:229-249
#title Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.
#cross-references MUID:88124198
#accession S04807
#status translation not shown
#molecule_type DNA
#residues 1-148 #label LAZ
#cross-references EMBL:Y00514; NID:g59365; PID:g59369
SUMMARY #length 148 #molecular-weight 17214 #checksum 3101

Query Match 5.0%; Score 119; DB 2; Length 148;
Best Local Similarity 26.7%; Pred. No. 5.62e-04;

Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;

Db 12 RSRKQSL-YIVGPTRTGKSTWARSIGVHN--YWNQND--WSSYNEDAIYNIVDDIPFKF 66
QY 163 RDWKTNVHFIVGPGCGSKAANFANPETTYWPKPKKWDGYHGERV-VVIDDF-YGW 220
Db 67 CPCWKQLVG-CORDFIVNPKYKGGKKVOKKSKPTIILANSDE-DW 109
QY 221 LP-WDDLLRLCDR-YPLTVK-TKGTVPFLARSILITSNQTPLEW 262

RESULT 5
ENTRY D42452 #type complete
TITLE C1 protein - tobacco yellow dwarf virus (strain Australia)
ORGANISM #formal_name tobacco yellow dwarf virus
DATE 15-Jan-1993 #sequence_revision 15-Jan-1993 #text_change 17-Mar-1999
ACCESSIONS D42452
REFERENCE A42452
#authors Morris, B.A.M.; Richardson, K.A.; Haley, A.; Zhan, X.; Thomas, J.E.
#journal Virology (1992) 187:633-642
#title The nucleotide sequence of the infectious cloned DNA component of tobacco yellow dwarf virus reveals features of geminiviruses infecting monocotyledonous plants.

#cross-references MUID:92188538
#accession D42452
#molecule_type DNA
#residues 1-295 #label MOR
#cross-references GB:M81103; NID:g335283; PID:g335287
SUMMARY #length 295 #molecular-weight 33651 #checksum 8742

Query Match 4.9%; Score 116; DB 2; Length 295;
Best Local Similarity 28.7%; Pred. No. 1.61e-03;
Matches 25; Conservative 29; Mismatches 26; Indels 7; Gaps 7;

Db 26 SSSAENLRDFLWD-KLSRFAIFFTAIAITELHODG-TPHLHCLIQDKRSINRDPFFDLE 83
QY 22 NNPSEDERKKIRLPISLFD-YFI-VGEEGNEGRTPHLOGFANFVKQTGNKVKY-L- 77
Db 84 GNHPNIOPAKNSQ-VLEYISKDGNVI 109
QY 78 GARCHIEKAKGTDQNKVCSKEGNLL 104

RESULT 6
ENTRY C71416 #type complete
TITLE Probable ABC transporter - Arabidopsis thaliana
ORGANISM #formal_name Arabidopsis thaliana #common_name mouse-ear cress
DATE 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 03-Dec-1998
ACCESSIONS C71416
REFERENCE A71400
#authors Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; Dirke, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson, S.A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Gielen, J.; Villarroel, R.; De Clerck, R.; Van Montagu, M.; Leclercq, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao, N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer, M.; Funk, B.; Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Puigdomenech, P.; Douka, A.; Vouklatou, E.; Milioni, D.; Hatzipoulos, P.; Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.; Moeres, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Reckman, S.; Ausage, W.; Cooke, R.; Berger, C.; Delsen, M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.; Schueller, C.; Chalwatzis, N.
#journal Nature (1998) 391:485-488
#title Analysis of 1.9 Mb of contiguous sequence from chromosome 4

```

```

of Arabidopsis thaliana
#cross-references MUID:98121113
#accession C71416
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-290 #label BEV
#cross-references GB:297338; NID:G2244870; PID:G326914; PID:G2244880
GENETICS
#map_position 4COP9-4G3845
SUMMARY
#length 290 #molecular-weight 32278 #checksum 8580
Query Match 4.3%; Score 102; DB 2; Length 290;
Best Local Similarity 36.1%; Pred. No. 1.78e-01;
Matches 13; Conservative 15; Mismatches 6; Indels 2; Gaps 2;

Db 178 LKGVSGIVR-PGVRQRCWFSCRMILLIGPPCGCKT 212
QY 147 FRGLAELKVGSKMRDWN-KTNVHFIVGPGCGKS 181

RESULT 7
ENTRY
TITLE genome polyprotein B - red clover mottle virus (strain S)
CONTAINS 24K viral proteinase; 32K proteinase cofactor; 58K membrane-binding protein; RNA-directed RNA polymerase (EC 2.7.7.48); VPg protein
ORGANISM #formal_name red clover mottle virus
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997
ACCESSION JQ1657; JQ0572; S19926
REFERENCE JQ1657
#authors Shanks, M.; Lomonosoff, G.P.
#journal J. Gen. Virol. (1992) 73:2473-2477
#title The nucleotide sequence of red clover mottle virus bottom component RNA.
#accession JQ1657
#molecule_type genomic RNA
#residues 1-1864 #label SHA
#cross-references EMBL:X64886; NID:G61463; PID:G61464
REFERENCE JQ0572
#authors Shanks, M.; Lomonosoff, G.P.
#journal J. Gen. Virol. (1990) 71:735-738
#title The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comovirus proteases.
#accession JQ0572
#molecule_type genomic RNA
#residues 912-1167 #label SH2
GENETICS
#map_position segment B
CLASSIFICATION #superfamily cowpea aphid-borne mosaic virus genome polyprotein B
KEYWORDS genome-linked protein; hydrolase; membrane protein; nucleotidyltransferase; polyprotein; proteinase
FEATURE
2-315 #product 32K proteinase cofactor #status predicted
#label PCF\
316-915 #product 58K membrane-binding protein #status predicted
#label MBP\
916-943 #product genome-linked protein VPg #status predicted
#label VPG\
944-1151 #product 24K viral proteinase #status predicted #label VPT\
1152-1864 #product RNA-directed RNA polymerase #status predicted #label RRP
SUMMARY
#length 1864 #molecular-weight 210254 #checksum 268
Query Match 4.2%; Score 99; DB 1; Length 1864;
Best Local Similarity 40.7%; Pred. No. 4.61e-01;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 509 DTYSRNTTTHWSGVRPPIVVDDF 535
of Arabidopsis thaliana
#cross-references MUID:98121113
#accession C71416
#status preliminary; nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-290 #label BEV
#cross-references GB:297338; NID:G2244870; PID:G326914; PID:G2244880
GENETICS
#map_position 4COP9-4G3845
SUMMARY
#length 290 #molecular-weight 32278 #checksum 8580
Query Match 4.3%; Score 102; DB 2; Length 290;
Best Local Similarity 36.1%; Pred. No. 1.78e-01;
Matches 13; Conservative 15; Mismatches 6; Indels 2; Gaps 2;

Db 178 LKGVSGIVR-PGVRQRCWFSCRMILLIGPPCGCKT 212
QY 147 FRGLAELKVGSKMRDWN-KTNVHFIVGPGCGKS 181

RESULT 7
ENTRY
TITLE genome polyprotein B - red clover mottle virus (strain S)
CONTAINS 24K viral proteinase; 32K proteinase cofactor; 58K membrane-binding protein; RNA-directed RNA polymerase (EC 2.7.7.48); VPg protein
ORGANISM #formal_name red clover mottle virus
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Sep-1997
ACCESSION JQ1657; JQ0572; S19926
REFERENCE JQ1657
#authors Shanks, M.; Lomonosoff, G.P.
#journal J. Gen. Virol. (1992) 73:2473-2477
#title The nucleotide sequence of red clover mottle virus bottom component RNA.
#accession JQ1657
#molecule_type genomic RNA
#residues 1-1864 #label SHA
#cross-references EMBL:X64886; NID:G61463; PID:G61464
REFERENCE JQ0572
#authors Shanks, M.; Lomonosoff, G.P.
#journal J. Gen. Virol. (1990) 71:735-738
#title The primary structure of the 24K protease from red clover mottle virus: implications for the mode of action of comovirus proteases.
#accession JQ0572
#molecule_type genomic RNA
#residues 912-1167 #label SH2
GENETICS
#map_position segment B
CLASSIFICATION #superfamily cowpea aphid-borne mosaic virus genome polyprotein B
KEYWORDS genome-linked protein; hydrolase; membrane protein; nucleotidyltransferase; polyprotein; proteinase
FEATURE
2-315 #product 32K proteinase cofactor #status predicted
#label PCF\
316-915 #product 58K membrane-binding protein #status predicted
#label MBP\
916-943 #product genome-linked protein VPg #status predicted
#label VPG\
944-1151 #product 24K viral proteinase #status predicted #label VPT\
1152-1864 #product RNA-directed RNA polymerase #status predicted #label RRP
SUMMARY
#length 1864 #molecular-weight 210254 #checksum 268
Query Match 4.2%; Score 99; DB 1; Length 1864;
Best Local Similarity 40.7%; Pred. No. 4.61e-01;
Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Db 509 DTYSRNTTTHWSGVRPPIVVDDF 535

```

```
#cross-references MUID:98139904
#accession JE0170
##molecule_type mRNA
##residues 1-244 #label SIL
##cross-references GB:AF036875
GENETICS
#gene Mcg18
CLASSIFICATION #superfamily dnaJ amino-terminal homology
FEATURE
41-103
#domain dnaJ amino-terminal homology #label DNJ
SUMMARY
#length 244 #molecular-weight 28206 #checksum 9987
Query Match 4.1%; Score 97; DB 2; Length 244;
Best Local Similarity 21.5%; Pred. No. 8.61e-01;
Matches 14; Conservative 18; Mismatches 31; Indels 2; Gaps 2;
Db 79 FVELNEARYVLSRESRNYDHLH-SASPPKSSGSAEPKRYTOQTTHSWPPNAQYWAQ 137
QY 147 FRGLAELLKVGKMQ-RDKMTNVHFIIVGPGCKSKWAANFANPETIYWKPKNKWDG 205
Db 138 FHSVR 142
QY 206 YHGEK 210
RESULT 10
ENTRY S04125 #type complete
TITLE chlorophyll a/b-binding protein type III precursor - tomato
ORGANISM #formal_name Lycopersicon esculentum #common_name tomato
DATE 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change
21-Aug-1998
ACCESSIONS S04125
REFERENCE #authors Pichersky, E.; Brock, T.G.; Nguyen, D.; Hoffman, N.E.;
Piechulla, B.; Tanksley, S.D.; Green, B.R.
#journal Plant Mol. Biol. (1989) 12:257-270
#title A new member of the CAB gene family: structure, expression
and chromosomal location of cab-8, the tomato gene encoding
the type III chlorophyll a/b-binding polypeptide of
photosystem I.
#accession S04125
##molecule_type DNA
##residues 1-273 #label PIC
##cross-references EMBL:X15258
GENETICS
#gene cab-8
#map_position 10
#introns 48/3; 69/2
CLASSIFICATION #superfamily chlorophyll a/b-binding protein
KEYWORDS chloroplast; light-harvesting complex; photosystem I;
transmembrane protein
SUMMARY #length 273 #molecular-weight 29293 #checksum 7937
Query Match 4.1%; Score 96; DB 2; Length 273;
Best Local Similarity 22.8%; Pred. No. 1.17e+00;
Matches 13; Conservative 22; Mismatches 16; Indels 6; Gaps 5;
Db 95 WLAYGEVIN--GRFAM-LGAAGATAPEILGKAGLI-POETALAWFOFGVPPACT-Y 146
QY 220 WLPWDDLLRLCDRYPLTVKTKGTVP-FLARSILITSNQIPLEWYSSTAVPAVEALY 275
RESULT 11
ENTRY A64380 #type complete
TITLE phosphoglycerate kinase (EC 2.7.2.3) - Methanococcus
ORGANISM jannaschii
#formal_name Methanococcus jannaschii
DATE 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change
10-Oct-1997
ACCESSIONS A64380
REFERENCE #authors Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann,
R.D.; Sutton, G.G.; Blake, J.A.; FitzGerald, L.M.; Clayton,
R.A.; Gocayne, J.D.; Kerlavage, A.R.; Dougherty, B.A.;
Tombl, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.;
Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.;
Scott, J.L.; Geoghegan, N.S.M.; Weidman, J.F.; Fuhrmann,
J.D.; Nguyen, D.; Utterback, T.R.; Kelley, J.M.; Peterson,
J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts,
K.M.; Hurst, M.A.; Kaine, B.P.; Borodovsky, M.; Klenk,
H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.
#journal Science (1996) 273:1058-1073
#title Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii.
#cross-references MUID:96337999
#accession A64380
#status preliminary; nucleic acid sequence not shown;
translation not shown
##molecule_type DNA
##residues 1-417 #label BUL
##cross-references GB:U67512; GB:L77117; NID:gl591352; PID:gl5922299;
TIGR:MJ0641; PID:gl510724
GENETICS
#map_position REV571288-570035
CLASSIFICATION #superfamily phosphoglycerate kinase
KEYWORDS ATP; phosphotransferase
SUMMARY #length 417 #molecular-weight 46133 #checksum 9304
Query Match 4.1%; Score 96; DB 2; Length 417;
Best Local Similarity 30.8%; Pred. No. 1.17e+00;
Matches 16; Conservative 12; Mismatches 20; Indels 4; Gaps 4;
Db 130 KKWE-NITPKKQAEINLKRLA-PLFDYF-VNDAFAAAHRAQPSLVGFSYVM 178
QY 15 KRWVTLNPNSEDERKKIRELPISLFDYFVIGEGNEGRT-PHLOGFANFV 65
RESULT 12
ENTRY T00490 #type complete
TITLE nonstructural protein precursor - himetobi P virus
ORGANISM #formal_name himetobi P virus
DATE 29-Jan-1999 #sequence_revision 29-Jan-1999 #text_change
29-Jan-1999
ACCESSIONS T00490
REFERENCE #authors Nakashima, N.; Sasaki, J.
#submission submitted to the EMBL Data Library, August 1998
#description Complete nucleotide sequence of an insect picorna-like virus,
Himetobi P virus (HiPv).
#accession T00490
#status preliminary
#residues 1-1777 #label NAK
##cross-references EMBL:AB017037; NID:dl226972; PID:dl033516
SUMMARY #length 1777 #molecular-weight 201730 #checksum 5486
Query Match 4.1%; Score 96; DB 3; Length 1777;
Best Local Similarity 45.8%; Pred. No. 1.17e+00;
Matches 11; Conservative 5; Mismatches 8; Indels 0; Gaps 0;
Db 610 YSRASEQEFVDGVTGOLVTVFDDF 633
QY 194 YWKPPKKNWDGYGKRVVVIDF 217
RESULT 13
ENTRY H70654 #type complete
TITLE hypothetical protein Rv3851 - Mycobacterium tuberculosis
ORGANISM (strain H37RV)
#formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
17-Jul-1998
ACCESSIONS H70654
REFERENCE #authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,
C.; Harris, D.; Gordon, S.V.; Eiglmeier, K.; Gas, S.; Barry
III, C.E.; Tekala, F.; Badcock, K.; Basham, D.; Brown, D.;
```

```

REFERENCE.
#authors Li, L.A.; Gibson, J.L.; Tabita, F.R.
#journal Plant Mol. Biol. (1993) 21:753-764
#title The Rubisco activase (rca) gene is located downstream from
        rbcS in Anabaena sp. strain CA and is detected in other
        Anabaena/Nostoc strains.
#cross-references MUID:93222475
#accession S33627
#molecule_type DNA
#residues 1-415 #label LIL
#cross-references EMBL:X67942; NID:g296413; PID:g296414
GENETICS
#gene rca
KEYWORDS
ATP
SUMMARY
length 415 #molecular-weight 46594 #checksum 8862

Query Match 4.0%; Score 95; DB 2; Length 415;
Best Local Similarity 34.9%; Pred. No. 1.59e+00;
Matches 15; Conservative 10; Mismatches 16; Indels 2; Gaps

```

```

      matches  15;  conservative  10;  mismatches  10;  indels  2;  gaps
Db  54 MGVEV-TLISGELESPOAGDPARILRLRYRETAELIKVGRKM 95
      ::  |  ||  :  |  |  :  |  :  :  :  |  ||  :  ||  ||
QY  119 LSTAVSTLESGLIVTVAEQHPVTEVK-NFRGLAELLKVSGRM 160

```

Search completed: Wed Dec 22 10:12:27 1999
Job time.: 20 secs.

WATERMAN

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:12:44 1999; MasPar time 11.43 Seconds
Tabular output not generated. 776.734 Million cell updates/sec

Title: >US-09-209-961-3
Description: (1-314) from US09209961.pep
Perfect Score: 2358
Sequence: 1 MPSKNGRSGPQPHRWFT.....GGOFVTLSPPCPEFYINY 314

Scoring table: PAM 150
Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: swiss-prot37
1:swissprot

Statistics: Mean 47.377; Variance 72.055; scale 0.658

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	119	5.0	148	1 Y18K_MSVS	HYPOTHETICAL 17.2 KD P	3.62e-05
2	119	5.0	153	1 Y18K_MSVN	HYPOTHETICAL 17.7 KD P	3.62e-05
3	116	4.9	295	1 VAL1_TYDVA	ALI PROTEIN (C1 PROTEI	1.22e-04
4	99	4.2	1864	1 VGNB_CPMV	GENOME POLYPROTEIN B I	8.22e-02
5	100	4.2	2157	1 POLG_HRV1B	GENOME POLYPROTEIN [CO	5.72e-02
6	96	4.1	273	1 CB13_LYCES	CHLOROPHYLL A-B BINDIN	2.39e-01
7	96	4.1	417	1 PKG_METJA	PHOSPHOGLYCERATE KINA	2.39e-01
8	94	4.0	503	1 AMP_P_RAT	ALPHA-AMYLASE PANCREAT	4.81e-01
9	95	4.0	2153	1 POLG_HRV16	GENOME POLYPROTEIN [CO	3.40e-01
10	91	3.9	136	1 Y16K_CSMV	HYPOTHETICAL 15.8 KD P	1.34e+00
11	91	3.9	159	1 REEC_ECOLI	SIGMA-E FACTOR REGULAT	1.34e+00
12	91	3.9	211	1 G12_ANOGA	PROTEIN G12 PRECURSOR	1.34e+00
13	91	3.9	243	1 CLCA_BOVIN	CLATHRIN LIGHT CHAIN A	1.34e+00
14	91	3.9	248	1 CLCA_HUMAN	CLATHRIN LIGHT CHAIN A	1.34e+00
15	91	3.9	248	1 CLCA_RAT	CLATHRIN LIGHT CHAIN A	1.34e+00
16	92	3.9	313	1 Y01H_MYCTU	HYPOTHETICAL 32.9 KD P	9.55e-01
17	91	3.9	376	1 MFC1_CITFR	MODIFICATION METHYLASE	1.34e+00
18	91	3.9	590	1 ILVB_PORPU	ACETOLACTATE SYNTHASE	1.34e+00
19	92	3.9	812	1 FLGD_ECOLI	OUTER MEMBRANE USHER P	9.55e-01
20	91	3.9	820	1 ALPHA-GLUCA	1,4-ALPHA-GLUCAN BRANC	1.34e+00
21	92	3.9	856	1 POLG_PVYO	GENOME POLYPROTEIN [CO	9.55e-01
22	93	3.9	879	1 GUNI_CLOTM	ENDOGLUCANASE I PRECUR	6.79e-01
23	91	3.9	1679	1 YMF9_YEAST	HYPOTHETICAL 187.1 KD	1.34e+00

24 92 3.9 1875 1 ITB4_HUMAN INTEGRIN BETA-4 SUBUNI 9.55e-01
25 92 3.9 2150 1 POLG_HRV2 GENOME POLYPROTEIN [CO 9.55e-01
26 92 3.9 2226 1 POLG_HPAV4 GENOME POLYPROTEIN [CO 9.55e-01
27 92 3.9 2226 1 POLG_HPAV2 GENOME POLYPROTEIN [CO 9.55e-01
28 92 3.9 2226 1 POLG_HPAV8 GENOME POLYPROTEIN [CO 9.55e-01
29 92 3.9 2227 1 POLG_HPAVM GENOME POLYPROTEIN [CO 9.55e-01
30 92 3.9 2227 1 POLG_HPAVL GENOME POLYPROTEIN [CO 9.55e-01
31 92 3.9 2227 1 POLG_HPAVH GENOME POLYPROTEIN [CO 9.55e-01
32 92 3.9 2230 1 POLG_HPAVS GENOME POLYPROTEIN [CO 9.55e-01
33 99 3.8 456 1 NY5R_RAT NEUROPEPTIDE Y RECEPTO 2.60e+00
34 99 3.8 466 1 NY5R_MOUSE NEUROPEPTIDE Y RECEPTO 2.60e+00
35 89 3.8 781 1 GCS1_CAEEL PROBABLE MANNOSYL-OLIG 2.60e+00
36 90 3.8 1130 1 C2TA_HUMAN MHC CLASS II TRANSACTI 1.87e+00
37 89 3.8 1866 1 VGNB_CPMV GENOME POLYPROTEIN B I 2.60e+00
38 90 3.8 3859 1 RPOA_LELV POL POLYPROTEIN (ORF1A 1.87e+00
39 90 3.8 4563 1 APB_HUMAN APOLIPOPROTEIN B-100 P 1.87e+00
40 88 3.7 137 1 MGSB_BACSU METHYLGLOXAL SYNTHASE 3.61e+00
41 87 3.7 284 1 KDSA_ECOLI 2-DEHYDRO-3-DEOXYPHOS 5.00e+00
42 88 3.7 455 1 NY5R_HUMAN NEUROPEPTIDE Y RECEPTO 3.61e+00
43 88 3.7 497 1 KPYC_ARATH PROBABLE PYRUVATE KINA 3.61e+00
44 88 3.7 718 1 STT3_YEAST OLIGOSACCHARYL TRANSFE 3.61e+00
45 87 3.7 1699 1 POLN_LORDV NON-STRUCTURAL POLYPRO 5.00e+00

ALIGNMENTS

REF:ILT 1
ID Y18K_MSVS STANDARD; PRT; 148 AA.
AC P14989;
DT 01-APR-1990 (REL. 14, CREATED)
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 17.2 KD PROTEIN.
OS MAIZE STREAK VIRUS (SOUTH-AFRICAN ISOLATE) (MSV).
OC VIRUSES; SSRNA VIRUSES; GEMINIVIRIDAE; MASTREVIROUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 86124198.
ZA LAZAROWITZ S.G.;
RT "Infectivity and complete nucleotide sequence of the genome of a South African isolate of maize streak virus.";
FL NUCLEIC ACIDS RES. 16:229-249(1988).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
EMBL; Y00514; G59369;
PIR; S04807; S04807.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 148 AA; 17214 MW; E0133DAF CRC32;

Query Match 5.0%; Score 119; DB 1; Length 148;
Best Local Similarity 26.7%; Pred. No. 3.62e-05;
Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;
Db 12 RSRKSL-YIVGPTRTGKSTWARSGLVHN--YWNQNV--WSSYNEDAIYIVDDIPKFK 66
QY 163 RDKKNVHEIVGPPCGCKSKWAANFANPTTWKPKKNWDYHGKVK-VVIDDF-YGW 220
Db 67 CPCWQLVG-CQDRITVNPYKGGKKKKVKQKSKPTIILANSDE-DW 109
CY 221 LP-WDDLRLCDR-YPLTVK-TKGGIVPFLARSILITSNOTPLEM 262

RESULT 2
ID Y18K_MSVN STANDARD; PRT; 153 AA.
AC P14978;
DT 01-APR-1990 (REL. 14, CREATED)

DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 17.7 KD PROTEIN.
 OS MAIZE STREAK VIRUS (NIGERIAN ISOLATE) (MSV).
 OC VIRUSES; SSDNA VIRUSES; GEMINIVIRIDAE; MASTREVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 85126910.
 RA MULLINEAUX P.M., DONSON J., MORRIS-KRSINICH B.A.M., BOULTON M.I.,
 RA DAVIES J.W.;
 RT "The nucleotide sequence of maize streak virus DNA."
 RL EMBO J. 3:3063-3068(1984).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X01633; G59362; -
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 153 AA; 17768 MW; FLE8189 CRC32;

 Query Match 5.08; Score 119; DB 1; Length 153;
 Best Local Similarity 26.7%; Pred. No. 3.62e-05;
 Matches 28; Conservative 28; Mismatches 37; Indels 12; Gaps 10;

 Db 12 RSKQSL-YVGTGRTGKSTWARSGLVHN--YQNNVD--WSSYNEDAIYVDDIPK 66
 QY 163 RDKWTNHFIVGPGCGSKAAANFANPETTYWPKNKNWGDYHGKV-VWIDDF-YGW 220

 Db 67 CPCKVLG-CQRFDFIVNPKYKKKKYQKSKPTIILANSDE-DW 109
 QY 221 LP-WDDLRLCDR-YPLTVK-TGGTVFPFLARSILITSNTPLEW 262

 RESULT 3
 ID VAL1.TYDVA STANDARD; PRT; 295 AA.
 AC P31617;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DE 01-OCT-1993 (REL. 27, LAST ANNOTATION UPDATE)
 DE AL1 PROTEIN (C1 PROTEIN).
 GN C1.
 OS TOBACCO YELLOW DWARF VIRUS (STRAIN AUSTRALIA) (TYDV).
 OC VIRUSES; SSDNA VIRUSES; GEMINIVIRIDAE; MASTREVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 92189538.
 RA MORRIS B.A.M., RICHARDSON K.A., HALEY A., ZHAN X., THOMAS J.E.;
 RT "The nucleotide sequence of the infectious cloned DNA component of
 RT tobacco yellow dwarf virus reveals features of geminiviruses
 RT infecting monocotyledonous plants."
 RL VIROLOGY 187:633-642(1992).
 CC -1- FUNCTION: MAY BE INVOLVED IN REPLICATION.
 CC -1- SIMILARITY: BELONGS TO GEMINIVIRUSES ALL1 PROTEIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M81103; G335287; -
 DR PIR; D42452; D42452.
 DR PFAM; PF00799; Geminivirus All; 1.
 SQ SEQUENCE 295 AA; 33651 MW; 520E4B9 CRC32;

 Query Match 4.98; Score 116; DB 1; Length 295;

- Best Local Similarity 28.7%; Pred. No. 1.22e-04;
 Matches 25; Conservative 29; Mismatches 26; Indels 7; Gaps 7;

 Db 26 SSSAENLRFLWD-KLSRFAIFFIAIATELHQDQ-TPHLHCLIQDKRSNRDPSFFDLE 83
 QY 22 NNPSEDERKKIRLEPISLFD-YFI-VGEEGNEGRTPHLQGFANFVKQTKENKVKWY-L- 77

 Db 84 GNHPNIQPAKNSQ-VLEYISKDGNVI 109
 QY 78 GARCHIEKAGTDOONKEYCSKEGNLL 104

 RESULT 4
 ID VGNB.RCMV STANDARD; PRT; 1864 AA.
 AC P35930;
 DT 01-JUN-1994 (REL. 29, CREATED)
 DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
 DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN B (CONTAINS: PROTEASE COFACTOR; MEMBRANE BINDING
 DE PROTEIN; VPG; PROTEASE (EC 3.4.22.-); RNA POLYMERASE (EC 2.7.7.48)).
 OS RED CLOVER MOTTLE VIRUS (RCMV).
 OC VIRUSES; SSNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; COMOVIRIDAE;
 CC COMOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-S;
 RX MEDLINE; 93019077.
 RA SHANKS M., LOMONOSOFF G.P.;
 RT "The nucleotide sequence of red clover mottle virus bottom component
 RT RNA."
 RL J. GEN. VIROL. 73:2473-2477(1992).
 CC -1- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
 CC -1- SIMILARITY: TO OTHER COMOVIRUSES GENOME POLYPROTEIN B.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X64886; G61464; -
 DR PIR; JQ1657; JQ1657.
 DR PFAM; PF00680; RNA_dep_RNA_pol; 1.
 DR PFAM; PF00910; RNA_helicase; 1.
 KW POLYPROTEIN; TRANSMEMBRANE; HYDROLASE; PROTEASE; TRANSFERASE;
 KW RNA-DIRECTED RNA POLYMERASE; ATP-BINDING.
 FT CHAIN 1 315
 FT CHAIN 316 915
 FT CHAIN 916 943
 FT CHAIN 944 1151
 FT CHAIN 1152 1864
 FT NP_BIND 483 490
 FT ACT_SITE 983 983
 FT ACT_SITE 1019 1019
 FT ACT_SITE 1109 1109
 SQ SEQUENCE 1864 AA; 210255 MW; 8F34B5BD CRC32;

 Query Match 4.28; Score 99; DB 1; Length 1864;
 Best Local Similarity 40.7%; Pred. No. 8.22e-02;
 Matches 11; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

 Db 509 DIVYSRNTTETHWSGYRROPVWIDDF 535
 QY 191 ETTYWKPKKNWGDYHGKRVWIDDF 217

 RESULT 5
 ID POLG.HRVLB STANDARD; PRT; 2157 AA.
 AC P12916; Q89704; Q82106; Q82107; Q82108; Q82109; Q82110; Q82111;
 DT 01-OCT-1989 (REL. 12, CREATED)
 DT 01-OCT-1989 (REL. 12, LAST SEQUENCE UPDATE)

[illegible]

RESULT	7	PGK METUA	STANDARD;	PRT;	417 AA.
ID	AC	Q58058;			
DY	DT	01-NOV-1997	(REL. 35, CREATED)		
LT	LT	01-NOV-1997	(REL. 35, LAST SEQUENCE UPDATE)		
DT	DT	01-NOV-1997	(REL. 35, LAST ANNOTATION UPDATE)		
DE	DE	PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).			
GN	GN	PKG OR M10641.			
OS	OS	METHANOCOCCUS JANNASCHII.			
CC	CC	ARCHAEA; EURYARCHAEOTA; METHANOCOCCALES; METHANOCOCCACEAE;			
OC	OC	METHANOCOCCUS.			
RN	RN	[1]			
RF	RF	SEQUENCE FROM N.A.			
KC	KC	STRAIN-JAL-1 / DSM	2661 / ATCC 43067;		
RX	RX	MEDLINE;	56337999.		
EA	EA	BULT C.J., WHITE O., OLSEN G.J., ZHOU L., FLEISCHMANN R.D.,			

X-RAY CRYSTALLOGRAPHY (2.15 ANGSTROMS) OF 2-853, AND REV. TO 547-548...
MEDLINE; 97238938.
RA HADFIELD A.T., LEE W.M., ZHAO R., OLIVEIRA M.A., MINOR I.,
RUECKERT R.R., ROSSMANN M.G.:
RT "The refined structure of human rhinovirus 16 at 2.15-A resolution:
RT implications for the viral life cycle.";
RT STRUCTURE 3:427-441(1997).
CC -!- FUNCTION: P3C POLYPEPTIDE IS A PROTEASE THAT CLEAVES AT CERTAIN
CC Q/Q SITES IN THE POLYPROTEIN. IT MAY BE A CYSTEINE PROTEASE.
CC -!- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC EACH OF WHICH IS COMPOSED OF ONE COPY EACH OF PROTEINS VP1, VP2,
CC VP3, AND VP4.
CC -!- PTM: SPECIFIC ENZYMIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: THE PROTEASE BELONGS TO PEPTIDASE FAMILY C3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR	EMBL; L24917; G409464; -.
DR	PDB; IAYN; 21-JAN-98.
DR	PDB; IAYM; 21-JAN-98.
DR	PFAM; PF00073; rhv; 3.
DR	PFAM; PF00548; Cys-protease-3C; 1.
DR	PFAM; PF00680; RNA_dep_RNA_pol; 1.
DR	PFAM; PF00910; RNA_helicase; 1.
DR	PFAM; PF00947; Pico_P2A; 1.
KW	POLYPROTEIN; COAT PROTEIN; TRANSFERASE;
KW	RNA-DIRECTED RNA POLYMERASE; HYDROLASE; THIOL PROTEASE; MYRISTYLATION;
KW	3D-STRUCTURE.
FT	CHAIN 2 69
FT	CHAIN 70 330
FT	CHAIN 331 568
FT	CHAIN 569 853
FT	CHAIN 854 995
FT	CHAIN 996 1090
FT	CHAIN 1091 1412
FT	CHAIN 1413 1499
FT	CHAIN 1490 1510
FT	CHAIN 1511 1693
FT	CHAIN 1694 2153
FT	CHAIN 2 2
FT	ACT_SITE 1657 1657
FT	ACT_SITE 1671 1671
FT	CONFLICT 547 548
SQ	SEQUENCE 2153 AA; 242242 MW; B81D1855 CRC32;

Query Match	4.0%	Score 95;	DB 1;	Length 2153;
Best Local Similarity	37.1%;	Pred. No. 3.40e-01;		
Matches 23;	Conservative	7;	Mismatches 28;	Indels 4;
				Gaps 3;

Db	1200	KQKRSEPVAMTHGPPGTGCKSTTSLARMINESDIYSLPDPKRYFDGYDQNSVIMD	1259
		: : :	
QY	160	MQRDKNTNVHFIVGPGCGKS--K-WAANFANPET-TYKKPPKNWDGYHGKEVVVID	215
		:	
Db	1260	DI	1261
		:	
QY	216	DF	217

RESULT	10	STANDARD;	PRT;	136 AA.
ID	Y16K_CSMW			
AC	PI8919;			
DT	01-NOV-1990	(REL. 16, CREATED)		
DT	01-NOV-1990	(REL. 16, LAST SEQUENCE UPDATE)		
DT	01-NOV-1990	(REL. 16, LAST ANNOTATION UPDATE)		
DE	HYPOTHETICAL 15.8 KD PROTEIN.			
OC	CHLORIS STRIATE MOSAIC VIRUS (CSMV).			
OC	VIRUSES: SSDNA VIRUSES: GEMINIVIRIDAE: MASTREVIRUS.			

[1] SEQUENCE FROM N.A.
MP MEDJINE; 58219532.
RX ANDERSEN H.T., RICHARDSON K.A., HARBISON S.A., MORRIS B.A.M.;
RA "Nucleotide sequence of the geminivirus chloris striate mosaic
RT virus.";
RL VIROLOGY 164:443-449(1988).
CC -1- WHILE THIS ORF DOES NOT HAVE AN ATG START CODON, THE SEQUENCE
CC SHOWS SIGNIFICANT HOMOLOGY, 69%, WITH THAT OF THE CORRESPONDING
CC ORF OF MAIZE STREAK VIRUS. ALSO, THE CORRESPONDING ORFS OF TWO
CC OTHER GEMINIVIRUSES DO NOT HAVE AN ATG START CODON.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M20021; -; NOT ANNOTATED_CDS.
DR PIR; JU0044;
DR JUP0044;
KW HYPOTHETICAL PROTEIN.
SO SEQUENCE 136 AA; 15762 MW; 5A046309 CRC32;

```
Query Match          3.9%; Score 91; DB 1; Length 136;
Best Local Similarity 26.2%; pred. No. 1.34e+00;
Matches            28; Conservative 44; Mismatches 44; Indels 11; Gaps 8;
```

RESULT 11
ID RSEC_ECOLI STANDARD; PRT; 159 AA.
AC P46187;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE THE SIGMA-A FACTOR REGULATORY PROTEIN RSEC.
EN ESEC.
OS ESCHERICHIA COLI.
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERiaceae;
OC ESCHERICHIA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
FA MISSILAS D.; RAINA S.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
FA CONNOLLY L., DE LAS PENAS A., GROSS C.A.;
RL SUBMITTED (OCT-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
FA NASHIMOTO H., SAITO N.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE: 97426617
RA BLATTNER F.R.; PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
RA GREGOR J., DAVIDSON N.W., KIRKPATRICK H.A., GOEDEN M.C., ROSE D.J.,
RA MAU B., SHAO Y.;
RT "the complete genomic sequence of Escherichia coli K-12".

RL SCIENCE 277:1453-1474(1997).
CC -1- FUNCTION: SEEMS TO MODULATE THE ACTIVITY OF RPOE (SIGMA-E).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
(POTENTIAL).
CC -1- SIMILARITY: TO H-INFLUENZAE HI589 AND HI850.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U37089; G1045631; -
DR EMBL; U37455; G1050877; -
DR EMBL; D64044; G987645; -
DR EMBL; AE000343; G1788923; -
DR ECGENE; EG13178; RSEC.
KW TRANSMEMBRANE; INNER MEMBRANE.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 101 121 POTENTIAL.
SQ SEQUENCE 159 AA; 16639 MW; E1CDF35 CRC32;
Query Match 3.9%; Score 91; DB 1; Length 159;
Best Local Similarity 35.5%; Pred. No. 1.34e+00;
Matches 11; Conservative 10; Mismatches 10; Indels 0; Gaps 0;
Db 51 IIVPCDEPLVPGKVGKVGIAEGSLSSALLV 81
QY 103 LLIECAPRSQGRSDUSTAVSLLESGLV 133
RESULT 12
ID G12.ANOGA STANDARD; PRT; 211 AA.
AC Q17040;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE PROTEIN G12 PRECURSOR (ANG12).
OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; NEMATOCERA; COLICOIDEA; COLICIDAE; ANOPHELES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUAKOKO;
RA MUELLER H.M., CRISANTI A.;
RL SUBMITTED (JUN-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
CC -1- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD
CC MEAL.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z22925; G311985; -
KW SIGNAL.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 211 PROTEIN G12.
SQ SEQUENCE 211 AA; 23568 MW; 9ED0DC5 CRC32;
Query Match 3.9%; Score 91; DB 1; Length 211;
Best Local Similarity 50.0%; Pred. No. 1.34e+00;
Matches 17; Conservative 4; Mismatches 13; Indels 0; Gaps 0;
Db 30 FDFVGLPLNDLLDAMRYLLTDKEVQVQLLYL 63
QY 214 IDDFYGLPWDDLLRLCDRLPLVTKRGTVFPL 247

RESULT 13
ID CICA_BOVIN STANDARD; PRT; 243 AA.
AC P04973;
DT 13-AUG-1987 (REL. 05, CREATED)
DT 13-AUG-1987 (REL. 05, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).
GN CLTA.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 87144634.
RA JACKSON A.P., SEOW H.-F., HOLMES N., DRICKAMER K., PARHAM P.;
RT "Clathrin light chains contain brain-specific insertion sequences and
a region of homology with intermediate filaments.";
RL NATURE 326:154-159(1987).
RN [2]
RP CLATHRIN HEAVY CHAIN-BINDING DOMAIN.
RX MEDLINE; 87144642.
RA BRODSKY F.M., GALLOWAY C.J., BLANK G.S., JACKSON A.P.,
SEOW H.-F., DRICKAMER K., PARHAM P.;
RT "Localization of clathrin light-chain sequences mediating heavy-chain
binding and coated vesicle diversity.";
RL NATURE 326:203-205(1987).
CC -1- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF
COATED PITS AND VESICLES.
CC -1- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3
HEAVY CHAINS AND 3 LIGHT CHAINS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND
VESICLES.
CC -1- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT
CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN
SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR
DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.
CC -1- THIS SEQUENCE IS IDENTICAL TO NON-BRAIN LC(A) EXCEPT THAT IT
CONTAINS A HYDROPHOBIC 30-RESIDUE, BRAIN-SPECIFIC INSERT.
CC THE NH2-TERMINAL PORTION OF THIS INSERT IS HOMOLOGOUS TO THE
BRAIN-SPECIFIC INSERT IN THE BRAIN LC(B) FORM.
CC -1- SIMILARITY: THERE IS A WEAK HOMOLOGY BETWEEN THE CENTRAL REGION OF
THE LIGHT CHAIN AND THE ALPHA-HELICAL PORTION OF VARIOUS IF
PROTEINS, PARTICULARLY CYTOKERATINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X04849; G508; -
DR EMBL; X04851; G510; -
DR PIR; A26599; A26599.
DR PIR; B26599; B26599.
DR PROSITE; PS00224; CLATHRIN_LIGHT_CHN_1; 1.
DR PROSITE; PS00581; CLATHRIN_LIGHT_CHN_2; 1.
DR PFAM; PF01086; Clathrin_lg_ch; 1.
DR MOD_RES 1 1 BLOCKED.
FT DOMAIN 95 157 INVOLVED IN BINDING CLATHRIN HEAVY
FT CHAIN.
FT VARSPIC 158 187 MISSING (IN NON BRAIN FORMS).
FT SIMILAR 121 146 TO INTERMEDIATE FILAMENTS.
FT CONFLICT 14 14 P -> H (IN LYMPHOCYTE LCA).
SQ SEQUENCE 243 AA; 26723 MW; 6C48315E CRC32;
Query Match 3.9%; Score 91; DB 1; Length 243;
Best Local Similarity 39.5%; Pred. No. 1.34e+00;
Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

Db 206 WERVARLDCDFNPKSSK-OAKDVSRM-RSVLISLKQAPL 241
 QY 223 WDDLRLCDRYPLTVKTKGGTVPFLARSILITSNOTPL 260

RESULT 14

ID CLCA_HUMAN STANDARD; PRT; 248 AA.
 AC P09496;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
 DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).
 GN CLTA.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 89034155.
 RA JACKSON A.P., PARHAM P.;
 RT "Structure of human clathrin light chains. Conservation of light chain polymorphism in three mammalian species.";
 RL J. BIOL. CHEM. 263:16688-16695(1988).
 CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF COATED PITS AND VESICLES.
 CC -!- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3 HEAVY CHAINS AND 3 LIGHT CHAINS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND VESICLES.
 CC -!- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M20471; G179397;
 DR EMBL; M20472; G307118;
 DR PIR; A30752; A30752.
 DR PIR; A30753; A30753.
 DR PIR; A31775; A31775.
 DR MM; I18960;
 DR PROSITE; PS00224; CLATHRIN_LIGHT_CHN_1; 1.
 DR PROSITE; PS00581; CLATHRIN_LIGHT_CHN_2; 1.
 DR PFAM; PF01086; Clathrin_lg_ch; 1.
 KW COATED PITS; ALTERNATIVE SPLICING; CALCIUM-BINDING.
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY CHAIN.
 FT VARSPLIC 163 192 MISSING (IN NON BRAIN FORMS).
 FT SIMILAR 126 151 TO INTERMEDIATE FILAMENTS.
 SQ SEQUENCE 248 AA; 27076 MW; 80CBA0B CRC32;

Query Match 3.9%; Score 91; DB 1; Length 248;
 Best Local Similarity 39.5%; Pred. No. 1.34e+00;
 Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

Db 211 WERVARLDCDFNPKSSK-OAKDVSRM-RSVLISLKQAPL 246
 QY 223 WDDLRLCDRYPLTVKTKGGTVPFLARSILITSNOTPL 260

RESULT 15

ID CLCA_RAT STANDARD; PRT; 248 AA.
 AC P08081;
 DT 01-AUG-1988 (REL. 08, CREATED)
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)

DE CLATHRIN LIGHT CHAIN A (BRAIN AND LYMPHOCYTE LCA).
 GN CLTA.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87178007.
 RA KIRCHHAUSEN T., SCARMATO P., HARRISON S.C., MONROE J.J., CHOW E.P., MATTALIANO R.J., RAMACHANDRAN K.L., SMART J.E., AHN A.H., BROSIUS J.;
 RT "Clathrin light chains LCA and LCB are similar, polymorphic, and share repeated heptad motifs.";
 RL SCIENCE 236:320-324(1987).
 CC -!- FUNCTION: CLATHRIN IS THE MAJOR PROTEIN OF THE POLYHEDRAL COAT OF COATED PITS AND VESICLES.
 CC -!- SUBUNIT: CLATHRIN COATS ARE FORMED FROM MOLECULES CONTAINING 3 HEAVY CHAINS AND 3 LIGHT CHAINS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC FACE OF COATED PITS AND VESICLES.
 CC -!- ALTERNATIVE PRODUCTS: MAMMALIAN CELLS CONTAIN 2 DISTINCT LIGHT CHAIN GENES, LC(A) AND LC(B), WHICH ENCODE PRODUCTS 60% RELATED IN SEQUENCE. TISSUE-SPECIFIC SPLICING PATTERNS GIVE RISE TO FOUR DISTINCT FORMS OF CLATHRIN LIGHT CHAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M15882; G203274;
 DR EMBL; M19261; G203278;
 DR EMBL; M19260; G203276;
 DR PIR; A25994; LRRYAL.
 DR PROSITE; PS00224; CLATHRIN_LIGHT_CHN_1; 1.
 DR PROSITE; PS00581; CLATHRIN_LIGHT_CHN_2; 1.
 DR PFAM; PF01086; Clathrin_lg_ch; 1.
 KW COATED PITS; ALTERNATIVE SPLICING; CALCIUM-BINDING.
 FT DOMAIN 100 162 INVOLVED IN BINDING CLATHRIN HEAVY CHAIN.
 FT VARSPLIC 163 192 MISSING (IN NON BRAIN FORMS).
 FT SIMILAR 126 151 TO INTERMEDIATE FILAMENTS.
 SQ SEQUENCE 248 AA; 26980 MW; 0E44BB68 CRC32;

Query Match 3.9%; Score 91; DB 1; Length 248;
 Best Local Similarity 39.5%; Pred. No. 1.34e+00;
 Matches 15; Conservative 10; Mismatches 11; Indels 2; Gaps 2;

Db 211 WERVARLDCDFNPKSSK-OAKDVSRM-RSVLISLKQAPL 246
 QY 223 WDDLRLCDRYPLTVKTKGGTVPFLARSILITSNOTPL 260

Search completed: Wed Dec 22 10:12:58 1999
 Job time : 14 secs.

W P S R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:15:55 1999; MasPar time 17.93 Seconds
Tabular output not generated. 276.392 Million cell updates/sec

Title: >US-09-209-961-5
Description: (1-233) from US09:09961.pep
Perfect Score: 1805
Sequence: 1 MTPYRRRRRRRRRSHLG.....RVMTYVQFRENLKDPPEP 233
Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 21266508 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 33.499; Variance 147.654; scale 0.227

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	111	6.1	143	11	R59940	rab Stage 1 VH constr
2	108	6.0	143	27	W31688	Transplanted MAB HPI/
3	108	6.0	143	13	R70253	Transplanted VH seque
4	108	6.0	143	10	R55208	Humanised HPI/2 heavy
5	108	6.0	143	36	W72432	Transplanted VH seque
6	104	5.8	325	33	W58645	Amino acid sequence o
7	103	5.7	124	11	R59930	SSE VH.
8	103	5.7	143	11	R59928	STAW VH.
9	103	5.7	143	11	R59941	Anti-VLA4 rab Stage 2
10	102	5.7	449	13	R65199	VPL protein.
11	103	5.7	657	6	R29580	PMR-1 gene product.
12	101	5.6	143	36	W72434	Humanised VH sequence
13	101	5.6	143	10	R55210	Humanised HPI/2 heavy
14	101	5.6	143	27	W31690	Humanised anti-VLA-4
15	101	5.6	143	11	R59932	AS VH.
16	101	5.6	143	11	R59751	Anti-VLA4 Ab AS VH se

17	101	5.6	143	13	R70255	Anti-VLA-4 humanized
18	99	5.5	143	11	R59931	KRS VH.
19	99	5.5	449	4	R38229	Chicken anaemia virus
20	98	5.4	449	16	R88497	VPI of chicken infect
21	97	5.4	497	30	W49912	Mouse CLK serine/thre
22	95	5.3	211	22	W06311	Hunt14.
23	96	5.3	316	9	R45924	Modified thermolysin-
24	95	5.3	402	26	R64200	Prostaglandin recepto
25	96	5.3	1485	25	W17887	Photorhabdus luminesc
26	96	5.3	1485	31	W56546	Toxin TscA, encoded b
27	93	5.2	117	1	P80149	Biosynthetic antibody
28	93	5.2	124	15	R82971	F85 humanized VH regi
29	94	5.2	163	25	W32556	Mycobacterium tubercu
30	94	5.2	163	35	W67296	Mycobacterium tubercu
31	94	5.2	163	37	W81059	M. tuberculosis immun
32	94	5.2	183	3	P60624	Sequence B encoded by
33	93	5.2	316	1	P91634	Sequence encoding hea
34	93	5.2	316	17	R97801	B.stearothermophilus
35	93	5.2	316	9	R49837	Thermolysin-like meta
36	93	5.2	316	26	W19127	Thermolysin protease
37	93	5.2	316	8	R40972	N227R mutant neutral
38	93	5.2	316	26	W19126	Thermolysin protease
39	93	5.2	316	17	R97802	B.stearothermophilus
40	93	5.2	316	17	R97795	B.stearothermophilus
41	93	5.2	316	17	R97796	B.stearothermophilus
42	93	5.2	316	8	R41515	Thermolysin like prot
43	93	5.2	377	39	W89556	Triticum sp. cysteine
44	94	5.2	518	39	W67616	A. nidulans phenylace
45	93	5.2	552	2	R08203	Structural gene produ

ALIGNMENTS

RESULT
ID R59940 standard; Protein; 143 AA.
AC R59940;
DE 09-MAR-1995 (first entry)
DE rab Stage 1 VH construct.
KW Humanised; human; mouse; recombinant; antibody; anti-VLA4; heavy;
KW light; variable; framework; rab; monoclonal; CDR; HPI/2; asthma; CDR;
KW complementarity determining region; inflammatory bowel disease.
OS Synthetic.
FH Key Location/Qualifiers
FT peptide 1..19 /note= "Signal peptide"
FT protein 20..143 /note= "Mature VH"
FT region 27..29 /note= "Murine residues"

W09416094-A.
PD 21-JUL-1994. U00266.
PF 07-JAN-1994; US-004798.
PR 12-JAN-1993; US-004798.
PE (BIO.) BIOGEN INC.
PI Carr FJ, Lobb RR, Tempest PR;
F2 WPI: 94-249240/30.
F3 N-PSDB; Q69901.
DR Recombinant and humanised anti-VLA4 antibodies - contg.
DT non-human CDRs in their light and heavy chains, used to treat
FT (non-)specific inflammation, e.g. asthma
PC Example 3: Page 90; 140pp; English.
CC This sequence represents the stage 1 heavy chain variable region (VH)
CC of the recombinant antibody (rab) of the invention. The VH sequence has
CC been CDR grafted and comprises a NEMO framework and CDRs derived from
CC the murine monoclonal antibody. HPI/2. In addition to CDR grafting, the
CC stage 1 VH construct contains selected framework changes. Just prior to
CC CDR1 a block of sequences have been changed to murine residues, Phe27,
CC Asn28, Ile29 and Lys30. Although these four residues are not nominally
CC included in CDR1, structurally they are included in part of the CDR loop,
CC and are empirically included as part of CDR1. Arg was also changed to
CC Asp at position 94. This sequence was used in the production of a
CC humanised recombinant anti-VLA4 antibody (rAb). The humanised rab
CC produced comprises non-human CDRs at VH positions 31-35 (CDR1), 50-65

CC (CDR2) and 95-102 (CDR3). It may further comprise non-human residues at framework positions 27-30, 75, 77-79 or 66-67, or 69-71 and 69-71, or 84-85, or 38 and 40 or 24. The VL of the rab has non-human CDRs at positions 24-34 (CDR1), 50-56 (CDR2) and 89-97 (CDR3). It may also have non-human residues at framework positions 60 and 67. The non-human CDRs are derived from HPI/2, which is a member of Kabat V with no unusual residues. The rab can be used to treat inflammation in mammals, esp. asthma or inflammatory bowel disease.

CC Sequence 143 AA;

Query Match 6.1%; Score 111; DB 11; Length 143;

Best Local Similarity 26.5%; Pred. No. 2.04e+00;

Matches 18; Conservative 21; Mismatches 25; Indels 4; Gaps 4;

Db 39 lctctvgfnikdtyhm-wvrgppgrglewgridpsgdktydkpfgvrvtmlvtdsknq 97

QY 167 LDSTIDYFQPNKRNQLWRLQTSQNVHDVG-LGTAFENSKYDQDYNIRVTMYVQ-FR-E 223

Db 98 fsrlsl 105

QY 224 FNLDKPPPL 231

RESULT 2

ID W31688 standard; Protein; 143 AA.

AC W31688;

DT 14-APR-1998 (first entry)

DE Transplanted MAB HPI/2 heavy chain variable region.

KW anti-VLA-4; monoclonal antibody; MAB; human; CD34+ cell; treatment;

KW cancer; VLA-4 antigen; chemotherapy.

OS Chimeric - Mus sp.

OS Chimeric - Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..19

FT Protein /note= "signal peptide"

FT Protein 20..143

FT Protein /note= "mature peptide"

PN US5695755-A.

PD 09-DEC-1997.

PF 05-JUN-1995; 463128.

PR 15-NOV-1993; US-436339.

PR 13-NOV-1992; US-977702.

PR 05-JUN-1995; US-463128.

PA (PAPA/) PAPAYANNOPOULOU T.

PI Papayannopoulou T;

DR N-PSDB; V02233.

DR N-PSDB; 98-041234/04.

PT Cancer treatment involving extracorporeal separation of CD34+ cells

PT - after peripheralisation by administering anti-VLA-4 antibody

PS Example 6; Columns 17-20; 28pp; English.

CC This is the transplanted Vh sequence of the anti-VLA-4 murine monoclonal

CC antibody HPI/2 heavy chain variable region in a human framework. This

CC is used in the preparation of a humanised anti-VLA-4 antibody. The

CC anti-VLA-4 antibodies can be used in a method for treating cancer in a

CC patient. The method comprises peripheralising CD34+ cells by

CC administering the anti-VLA-4 antibody. This antibody blocks the binding

CC of VLA-4 antigen on the surface of the CD34+ cells to VCAM or

CC fibronectin. Peripheral blood containing the CD34+ cells is collected

CC from the patient by leukapheresis. The CD34+ cells is enriched by

CC immunoadsorption using anti-CD34 antibodies. Chemotherapy and/or

CC radiotherapy can be administered to the patient and the enriched CD34+

CC cells is returned to the patient's circulating blood. The method is

CC especially used for restoring haematopoietic stem cells to patients who

CC have undergone myeloablative chemotherapy and radiotherapy.

CC Sequence 143 AA;

Query Match 6.0%; Score 108; DB 27; Length 143;

Best Local Similarity 26.6%; Pred. No. 3.40e+00;

Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;

Db 39 lctctvgfnikdtyhm-wvrgppgrglewgridpsgdktydkpfgvrvtmlvtdsknq 97

QY 167 LDSTIDYFQPNKRNQLWRLQTSQNVHDVG-LGTAFENSKYDQDYNIRVTMYVQ-FR-E 223

Db 98 fsrl 101

QY 224 FNLK 227

RESULT 3

ID R70253 standard; Protein; 143 AA.

AC R70253;

DT 20-OCT-1995 (first entry)

DE Transplanted Vh sequence.

KW VLA-4; reticulocyte; VCAM-1; endothelium; sickle cell anemia;

KW necrotizing enterocolitis; monoclonal antibody; MAB; heavy chain;

KW Vh; complementarity determining region; CDR; humanized antibody.

OS Mus sp.

FH Key Location/Qualifiers

FT Peptide 1..19

FT region /label= Sig_peptide

FT region 31..35

FT region /label= CDR1

FT region 50..66

FT region /label= CDR2

FT region 99..110

FT region /label= CDR3

PN W09507717-A.

PD 23-MAR-1995.

PF 15-SEP-1994; U10395.

PR 15-SEP-1993; US-122228.

PA (GEOR-) GEORGIA TECH RES CORP.

PA (UYEM-) UNIV EMORY.

PI Eckman JR, Swerlick RA, Wick TM;

DR WPI; 95-131188/17.

DR N-PSDB; Q83572.

PT Method for inhibiting binding between VLA-4 expressing

PT reticulocytes and VCAM-1 expressing endothelial cells - used as

PT a model for studying sickle cell anaemia

PS Disclosure; Page 29; 48pp; English.

CC Mouse anti-VLA-4 MAB HPI/2 VH and V-kappa regions, given in

CC R70253-54, respectively, were transplanted into Kabat frameworks

CC to obtain humanized HPI/2 heavy chain (R70255) and light chain

CC (R70256) constructs, which were used to produce humanized anti-VLA-4

CC antibody capable of blocking the binding of sickle erythrocytes to

CC activated endothelial cells.

CC Sequence 143 AA;

Query Match 6.0%; Score 108; DB 13; Length 143;

Best Local Similarity 26.6%; Pred. No. 3.40e+00;

Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;

Db 39 lctctvgfnikdtyhm-wvrgppgrglewgridpsgdktydkpfgvrvtmlvtdsknq 97

QY 167 LDSTIDYFQPNKRNQLWRLQTSQNVHDVG-LGTAFENSKYDQDYNIRVTMYVQ-FR-E 223

Db 98 fsrl 101

QY 224 FNLK 227

RESULT 4

ID R55208 standard; Protein; 143 AA.

AC R55208;

DT 24-JAN-1995 (first entry)

DE Humanised HPI/2 heavy chain variable region.

KW Variable region; heavy; light; chains; anti-VLA-4; murine; monocytes;

KW monoclonal antibody; HPI/2; VLA-4 complex; peripheral blood; B cells;

KW T cells; thymocytes; melanoma cells; marrow blast cells; haematopoiesis;

KW erythroblasts; peripheralisation; CD34 cells; blocking agent; cytokine;

KW VLA-4 surface antigen; bone marrow reconstitution; gene therapy;

KW autologous; homing; stem cells; cytokine-induced peripheralisation;

KW Kabat; NEMM; framework; REI; kappa; transplantation; M13 mutagenesis;

KW vectors.

OS Chimeric - Mus musculus,

OS Chimeric - Homo sapiens.


```

FH Key Location/Qualifiers
FT peptide 1..19
FT protein /note= "Signal peptide"
FT protein 20..143
FT peptide /note= "Humanised HPI/2"
FT peptide 46..49
FT peptide /note= "Murine residues"
FT peptide 50..54
FT peptide /label= CDR1
FT peptide 59..85
FT peptide /label= CDR2
FT misc_difference 113
FT peptide /note= "Murine residue"
FT peptide 118..129
FT peptide /label= CDR3
PN WO9411027-A.
PD 26-MAY-1994.
PP 15-NOV-1993; U11060.
PR 13-NOV-1992; US-977702.
PI (UNIW ) UNIV WASHINGTON.
PA Papayannopoulou T;
DR WPI; 94-183163/22.
DR N-PSDB; Q65624.
PT Peripheralisation of CD34 positive stem cells in peripheral blood
PT - using blocking agent for VLA-4 surface antigen, useful for
PT reconstituting bone marrow after AIDS or cancer treatment and in
PT gene therapy
PS Disclosure; Page 36-37: 52pp; English.
CC The sequences given in R55208-09 represent the humanised variable
CC regions of the heavy and light chains of anti-VLA-4 murine monoclonal
CC antibody HPI/2. The Kabat NEMO framework was used to accept the heavy
CC chain CDRs and the Kabat REI framework was chosen to accept the kappa
CC chain CDRs. Transplantation of the CDRs was achieved using M13
CC mutagenesis vectors and synthetic oligonucleotides containing the
CC CDR-encoding sequences flanked by short sequences derived from the
CC frameworks. The VH sequence, in addition to the CDRs, contains murine
CC amino acids at positions 27-30 and an Arg to Asp change at position 94.
CC The VL sequence contains no further murine residues. The VLA-4 complex
CC is expressed at substantial levels on normal peripheral blood B and T
CC cells, thymocytes, monocytes and some melanoma cells as well as on
CC marrow blast cells and erythroblasts. The humanised HPI/2 antibody,
CC may be used in the peripheralisation of CD34 cells by acting as a
CC blocking agent of VLA-4 surface antigen. The isolated cells are
CC useful for autologous bone marrow reconstitution and as carriers in
CC gene therapy of acquired and inherited diseases. The peripheral-
CC isation method can also be used as a model for studying haematopoiesis,
CC homing of stem cells and cytokine-induced peripheralisation. Treatment
CC with the blocking agent increases the number of CD34 cells in peripheral
CC blood more effectively than treatment with cytokine.
SQ Sequence 143 AA;

Query Match 6.0%; Score 108; DB 10; Length 143;
Best Local Similarity 26.6%; Pred. No. 3.40e+00;
Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;

Db 39 lctvsgfnikdymh-wvrqpggglewgridpasgdktdkpfqrvrtmldvtsknq 97
QY 167 LDSTIDYFQPNKRNQLWRLQTSNGVDHVG-LGTAFENSKYDQDYNIRVTMYVQ-FR-E 223
Db 98 fsrl 101
QY 224 FNLR 227

RESULT 5
ID W72432 standard; Protein; 143 AA.
AC W72432;
DT 22-DEC-1998 (first entry)
DE Transplanted VH sequence.
KW peripheralising; CD34+ cell; haematopoietic stem cell; VLA-4 antigen;
KW cell proliferation; cancer; AIDS; gene therapy; chemotherapy;
KW acquired immune deficiency syndrome; radiotherapy; peripheral blood;
KW anti-VLA-4 antibody.

```

```

CS Synthetic.
CS Mus sr.
PN US5824304-A.
PD 20-OCT-1998.
PP 05-JUN-1995; 463298.
PR 15-NOV-1993; US-436339.
PR 13-NOV-1992; US-977702.
PR 05-JUN-1995; US-463298.
PA (PAPA) PAPAYANNOPOULOU T.
PI Papa/annopoulou T;
DI WPI; 98-582496/49.
DI N-FSDB; V66801.
DR Treatment of acquired immune deficiency syndrome by chemotherapy
DR and/or radiotherapy - with restoration of haematopoietic stem cells
DR enriched from peripheral blood after treatment with anti-VLA-4
DR antibody
PS Example 6: Column 19-20: 28pp; English.
CC A method has been developed for treating acquired immune deficiency
CC syndrome (AIDS). The method comprises: (1) peripheralising CD34+ cells
CC by administering an anti-VLA-4 antibody that blocks the binding of
CC VLA-4 antigen on the surface of the CD34+ cells to VCAM or fibronectin;
CC (2) collecting peripheral blood containing the CD34+ cells by
CC leukapheresis; (3) enriching the CD34+ cells by immunoadsorption using
CC anti-CD34 antibodies; (4) administering chemotherapy and/or radiotherapy
CC to the patient; and (5) returning the enriched CD34+ cells to the
CC patient's circulating blood. Experiments on macaques and baboons
CC indicate that the method increases peripheral blood leukocyte counts by
CC a factor of 2 and CFU-GM levels can be increased by a factor of 8-40.
CC The present sequence represents a transplanted VH sequence which is
CC used in an example from the present invention.
SQ Sequence 143 AA;

Query Match 6.0%; Score 108; DB 36; Length 143;
Best Local Similarity 26.6%; Pred. No. 3.40e+00;
Matches 17; Conservative 19; Mismatches 24; Indels 4; Gaps 4;

Db 39 lctvsgfnikdymh-wvrqpggglewgridpasgdktdkpfqrvrtmldvtsknq 97
QY 167 LDSTIDYFQPNKRNQLWRLQTSNGVDHVG-LGTAFENSKYDQDYNIRVTMYVQ-FR-E 223
Db 98 fsrl 101
QY 224 FNLR 227

RESULT 6
ID W59645 standard; Protein; 325 AA.
AC W59645;
DT 12-OCT-1998 (first entry)
DE Amino acid sequence of the AS164.1 clone.
KW Clone AS164.1; secreted protein; human cell; immuno-modulator; PCR;
KW anti-tumour agent; tissue growth promoter; haemostatic agent; primer;
KW thrombolytic agent; molecular weight marker; genetic disorder;
KW antibody.
KW Homo sapiens.
PW WO9821332-A2.
PD 22-MAY-1998.
PR 14-NOV-1997; U20740.
PR 13-NOV-1997; US-969515.
PR 15-NOV-1996; US-749745.
PR 02-JUN-1997; US-667678.
PA (GENY ) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LS, Spaulding V, Treacy M;
DR WPI; 98-297936/26.
DR N-PSDB; V41444.
PT New nucleic acid encoding secreted protein from human cells -
PT potentially useful, e.g. as immuno-modulators, anti-tumour agents,
PT promoters of tissue growth, haemostatic and thrombolytic agents
PS Claim 16; Pages 66-67; 98pp; English.
CC This is the amino acid sequence of the clone AS164-1, used in the
CC method of the invention involving secreted proteins from human cells
CC potentially useful as immuno-modulators, anti-tumour agents, tissue

```


--||| || | : || : |||| : ||| : ||| : ||| : ||

Q7 4 PRRR-YARRHRPSHGLQILRRRPLWHPR-HRYWRRR-KNGIFNTR 48

RESULT 11

ID R29580 standard; Protein; 657 AA.

AC R29580;

DT 22-APR-1993 .(first entry)

DE FMR-1 gene product.

LW Fragile X disease; sex chromosome; X chromosome; X linked syndrome;

KW X linked retardation; X linked manic depression; TKCR;

KW Martin-Bell syndrome; CA polymorphisms; PCR analysis; ss.

OS Homo sapiens.

FH Key

FT cds 1..3766

FT /tag- a

ZN W09220825-A.

P3 26-NOV-1992.

PF 22-MAY-1992; U04447.

FR 24-MAY-1991; US-705490.

PR 29-AUG-1991; US-751891.

PA (BAYU) FAYLOR COLLEGE MEDICINE.

PA (BYEM-) UNIV EMORY SCHOOL.

PA MEDICINE.

I Caskey CT, Nelson DL, Oostra BA, Pieretti M, Warren ST;

DR WPI: 92-415801/50.

N-PSDB: Q31890.

DR Gene sequence, related probes and cosmid(s) - useful in

PT diagnosing fragile X syndrome, X-linked mental retardation, manic

PT depressor, and Martin Bell syndrome

PS Claim-14; Page 54; 75pp; English.

CC This sequence represents the "FMR-1 gene product. It was isolated

CC as detailed in Q31890. The level of expression of the FMR-1 gene

CC may be determined by determining the level of this protein. This is

CC useful when trying to detect fragile X syndrome, X-linked retardation

CC X-linked manic depression, TKCR and Martin-Bell syndrome.

SQ Sequence 657 AA;

CC may be determined by determining the level of this protein. This is
CC useful when trying to detect fragile X syndrome, X-linked retardation,
CC X-linked manic depression, TKCR and Martin-Bell syndrome.
SQ Sequence 657 AA;

```

LL      18 ***** 13
W       4 PRRYRRRHPSHLGQILRRRPWL 29
        IIII IIIII : : III :|
RESULT 12
ID      W72434 standard; Protein; 143 AA.
AC      W72434;
DT      22-DEC-1998 (first entry)
DE      Humanised VH sequence.
KW      peripheralising; CD34+ cell; haematopoietic stem cell; VLA-4 antigen;
KW      cell proliferation; cancer; AIDS; gene therapy; chemotherapy;
KW      acquired immune deficiency syndrome; radiotherapy; peripheral blood;
KW      anti-VLA-4 antibody.
OS      Synthetic.
OS      Homo sapiens.
OS      Mus sp.
PN      US5824304-A.
FD      20-OCT-1998.
PR      05-JUN-1995; 463298.
PR      15-NOV-1993; US-436339.
PR      13-NOV-1992; US-977702.
PR      05-JUN-1995; US-463298.
PA      (PAPA/) PAPAYANNOPOULOU T.
PI      Papayannopoulos T;
DR      WPI; 98-582496/49.
DK      N-PSDB: V66803.
FT      Treatment of acquired immune deficiency syndrome by chemotherapy
PT      and/or radiotherapy - with restoration of haematopoietic stem cells
PT      enriched from peripheral blood after treatment with anti-VLA-4
PT      antibody

```

DR WP1; 98-582496/749.
DK N-PSDB; V65803.
F: Treatment of acquired immune deficiency syndrome by chemotherapy
PT and/or radiotherapy - with restoration of haematopoietic stem cells
PT enriched from peripheral blood after treatment with anti-VLA-4
PT ant:bouy

Search completed: Wed Dec 22 10:16:18 1999
Job time : 23 secs.

```

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
      Distribution rights by Oxford Molecular Ltd

protein - protein database search, using Smith-Waterman algorithm

MPsrch_pp      Run on:      Wed Dec 22 10:17:49 1998;  Maspar time 5.91 Seconds
                  -----470.024 Million cell updates/sec
Tabular output not generated.

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Query		Score	Match	Length	DB	ID	Description	Pred. No.
	Match	Score							
1	108	6.0	143	2	US-08-436-	Sequence 4, Applicatio	2.26e+00		
2	108	6.0	143	1	US-08-463-	Sequence 4, Applicatio	2.26e+00		
3	108	6.0	143	2	US-08-463-	Sequence 4, Applicatio	2.26e+00		
4	102	5.7	449	2	US-08-911-	Sequence 3, Applicatio	5.96e+00		
5	102	5.7	449	2	US-08-489-	Sequence 3, Applicatio	5.96e+00		
6	101	5.6	143	2	US-08-436-	Sequence 8, Applicatio	7.00e+00		
7	101	5.6	143	2	US-08-463-	Sequence 8, Applicatio	7.00e+00		
8	101	5.6	143	1	US-08-463-	Sequence 8, Applicatio	7.00e+00		
9	100	5.5	449	1	US-07-917-	Sequence 2, Applicatio	8.21e+01		
10	95	5.3	416	2	US-08-698-	Sequence 5, Applicatio	8.18e+01		
11	96	5.3	1087	1	US-08-264-	Sequence 5, Applicatio	1.55e+01		
12	93	5.2	124	2	US-08-657-	Sequence 16, Applicati	2.47e+01		
13	94	5.2	183	4	5168049-4	Patent No. 5168049.	2.11e+01		
14	93	5.2	316	1	US-08-656-	Sequence 1, Applicatio	2.47e+01		
15	93	5.2	316	1	US-08-038-	Sequence 1, Applicatio	2.47e+01		
16	93	5.2	1257	1	US-08-340-	Sequence 49, Applicati	2.47e+01		
17	90	5.0	77	2	US-08-726-	Sequence 174, Applicat	3.93e+01		
18	90	5.0	432	2	US-08-933-	Sequence 47, Applicati	3.93e+01		
19	91	5.0	614	3	PCT-US95-0	Sequence 21, Applicati	3.37e+01		
20	91	5.0	2291	2	US-08-286-	Sequence 29, Applicati	3.37e+01		
21	88	4.9	124	2	US-08-657-	Sequence 17, Applicati	5.34e+01		
22	89	4.9	559	2	US-08-406-	Sequence 20, Applicati	4.58e+01		
23	89	4.9	1794	2	US-08-311-	Sequence 13, Applicati	4.58e+01		

[illegible]

XX	Sequence 1, Application US/08056349
DE	XX
CC	Sequence 1, Application US/08056349
CC	Patent No. 5728544
CC	GENERAL INFORMATION:
CC	APPLICANT:
CC	TITLE OF INVENTION: NOVEL PROTEIN
CC	NUMBER OF SEQUENCES: 7

XX	Sequence 1, Application US/08656349
DE	
XX	
CC	Sequence 1, Application US/08656349
CC	Patent No. 5728544
CC	GENERAL INFORMATION:
CC	APPLICANT:
CC	TITLE OF INVENTION: NOVEL PROTEASE II
CC	NUMBER OF SEQUENCES: 7
CC	

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/038,932B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-171479
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-237606
FILING DATE: 14-AUG-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 4-273413

CC FILING DATE: 18-SEP-1992
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-861-3000
CC TELEFAX: 202-638-4203
CC TELEX: 248453
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 316 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 316 AA; 34331 MW; 527255

Query Match 5.2%; Score 93; DB 1; Length 316;
Best Local Similarity 27.8%; Pred. No. 2.47e+01;
Matches 15; Conservative 15; Mismatches 21; Indels 3; Gaps 3;

D_b 33 NTRNGNGIETVDAKYR^TTT^TPGSI^TWADADNOF-EASVDAPAVDAHHVAGVTVDVYK 85

Qy 46 NTRLRTEGYTVK-RTTVTTPSNAVDMNREKIDDFVPPGGGTNK TSIPEFYR 97

Search completed: Wed Dec 22 10:17:59 1999
Job time : 10 secs.

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	158	8.8	118	2	S56117	spermatid-specific pr	8.95e-09
2	153	8.5	77	2	B40973	spermatid-specific pr	5.14e-08
3	146	8.1	78	2	A40973	spermatid-specific pr	5.73e-07
4	146	8.1	79	2	S56116	spermatid-specific pr	5.73e-07
5	126	7.0	58	2	S34045	protamine - North Ame	4.29e-04
6	120	6.6	1538	2	H70846	hypothetical glycine-	2.85e-03
7	113	6.3	52	2	PN0081	sperm chromatid prote	2.44e-02
8	113	6.3	57	2	A34356	protamine - Japanese	2.44e-02
9	113	6.3	61	2	S39425	protamine p1 - duckbi	2.44e-02
10	114	6.3	113	2	S66936	probable membrane pro	1.80e-02
11	110	6.1	126	2	S58321	probable membrane pro	5.99e-02
12	110	6.1	129	2	B6488	LSU ribosomal protein	5.99e-02
13	110	6.1	1460	1	EDB81F	immediate-early prote	5.99e-02
14	108	6.0	87	2	S00180	spermatid protein s1	1.08e-01
15	109	6.0	225	2	C71159	hypothetical protein	8.05e-02
16	107	5.9	238	2	A57198	splicing factor, argl	1.45e-01
17	106	5.9	284	2	S26437	hypothetical protein	1.94e-01
18	105	5.8	31	1	CLHRZA	protamine z - Atlanti	2.59e-01
19	105	5.8	31	1	CLHRZ	protamine z - Pacific	2.59e-01
20	105	5.8	56	2	C58213	protamine II - Americ	2.59e-01
21	105	5.8	65	1	GACH	protamine - chicken	2.59e-01
22	104	5.8	815	2	S67675	probable membrane pro	3.45e-01
23	102	5.7	58	2	S10755	protamine St2b - hors	6.10e-01

```

##residues 1-77 #label WOU
REFERENCE S14085
#authors Martin-Ponthieu, A.; Wouters-Tyrou, D.; Belaiche, D.;
#journal Sautiere, P.; Schindler, P.; van Dorsselaer, A.
#title Eur. J. Biochem. (1991) 195:611-619
#cross-references MUID:91153298
#accession S14086
#molecule_type protein
#residues 22-77 #label MAR
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; spermatogenesis
FEATURE 1-21
22-77 #domain signal sequence #status predicted #label SIG\
#product protamine variant Sp2 #status experimental
#label MAR
SUMMARY #length 77 #molecular-weight 10485 #checksum 2088
Query Match 8.5%; Score 153; DB 2; Length 77;
Best Local Similarity 55.3%; Pred. No. 5.14e-08;
Matches 21; Conservative 7; Mismatches 7; Indels 3; Gaps 3;
Db 43 Y-RRYRRRRRRRRR-R-SYRRYRRRR 77
QY 3 YPRRRRRRRRSHLQILRRPWLVPVPHRYRWRR 40
RESULT 3
ENTRY A40973 #type complete
TITLE spermatid-specific protein T1 precursor - common cuttlefish
ALTERNATE_NAMES arginine-rich protamine; testis-specific protein T1
ORGANISM #formal_name Sepia officinalis #common_name common cuttlefish
DATE 21-Apr-1992 #sequence_revision 21-Apr-1992 #text_change
13-Sep-1998
ACCESSIONS A40973; S14085
REFERENCE A40973
#authors Wouters-Tyrou, D.; Chartier-Harlin, M.C.; Martin-Ponthieu,
#journal A.; Bouillon, C.; Van Dorsselaer, A.; Sautiere, P.
#title J. Biol. Chem. (1991) 266:17388-17395
#cross-references MUID:91373359
#accession A40973
#molecule_type protein
#residues 1-78 #label WOU
REFERENCE S14085
#authors Martin-Ponthieu, A.; Wouters-Tyrou, D.; Belaiche, D.;
#journal Sautiere, P.; Schindler, P.; van Dorsselaer, A.
#title Eur. J. Biochem. (1991) 195:611-619
#cross-references MUID:91153298
#accession S14086
#molecule_type protein
#residues 22-78 #label MAR
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; spermatogenesis
FEATURE 1-21
22-78 #domain signal sequence #status predicted #label SIG\
#product protamine variant Sp1 #status experimental
#label MAR
SUMMARY #length 78 #molecular-weight 10631 #checksum 3554
Query Match 8.1%; Score 146; DB 2; Length 78;
Best Local Similarity 50.0%; Pred. No. 5.73e-07;
Matches 19; Conservative 8; Mismatches 9; Indels 2; Gaps 2;
Db 43 Y-RRYRRRRRRRRR-SRRYRRRRSRRYRRRR 78
QY 3 YPRRRRRRRRSHLQILRRPWLVPVPHRYRWRR 40

```

```

RESULT 4
ENTRY S56116 #type complete
TITLE spermatid-specific protein T1 - longfin squid
ORGANISM #formal_name Loligo pealeii #common_name longfin squid
DATE 27-Oct-1995 #sequence_revision 03-Nov-1995 #text_change
03-May-1996
ACCESSIONS S56116
REFERENCE S56116
#authors Wouters-Tyrou, D.; Martin-Ponthieu, A.; Ledoux-Andula, N.;
#journal Kouach, M.; Jaquinod, M.; Subirana, J.A.; Sautiere, P.
#title Biochem. J. (1995) 309:529-534
#cross-references MUID:91153298
#accession S56116
#status preliminary
#molecule_type protein
#residues 1-79 #label WOU
CLASSIFICATION #superfamily sperm histone
SUMMARY #length 79 #molecular-weight 10788 #checksum 5396
Query Match 8.1%; Score 146; DB 2; Length 79;
Best Local Similarity 50.0%; Pred. No. 5.73e-07;
Matches 19; Conservative 8; Mismatches 9; Indels 2; Gaps 2;
Db 44 Y-RRYRRRRRRRRR-SRRYRRRRSRRYRRRR 79
QY 3 YPRRRRRRRRSHLQILRRPWLVPVPHRYRWRR 40
RESULT 5
ENTRY S34045 #type complete
TITLE protamine - North American opossum
ORGANISM #formal_name Didelphis virginiana, Didelphis marsupialis
DATE 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change
08-Sep-1997
ACCESSIONS S34045
REFERENCE S34045
#authors Winkfein, R.J.; Nishikawa, S.; Connor, W.; Dixon, G.H.
#journal Eur. J. Biochem. (1993) 215:63-72
#title Characterization of a marsupial sperm protamine gene and its
transcripts from the North American opossum (Didelphis
marsupialis).
#cross-references MUID:93345500
#accession S34045
#status preliminary
#molecule_type DNA
#residues 1-58 #label WIN
#cross-references EXBL:X74044; NID:g407062; PID:g407063
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus
SUMMARY #length 58 #molecular-weight 7941 #checksum 3630
Query Match 7.0%; Score 136; DB 2; Length 58;
Best Local Similarity 50.0%; Pred. No. 4.29e-04;
Matches 19; Conservative 8; Mismatches 8; Indels 3; Gaps 3;
Db 3 YRRRSRSRSRYGRR-RRRSRRRRSR-RRRRRG 38
QY 7 YRRRRRRRSHLQILRRPWLVPVPHRYRWRRKNG 43
RESULT 6
ENTRY H70846 #type complete
TITLE hypothetical glycine-rich protein RV3345c - Mycobacterium
ORGANISM #formal_name Mycobacterium tuberculosis
DATE 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change
24-Jul-1998
ACCESSIONS H70846
REFERENCE H70846
#authors Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher,

```

```

08-Sep-1997
ACCESSION# A34356
REFERENCE# A34356
#authors J. Biol. Chem. (1989) 264:17627-17630
#journal J. Biol. Chem. (1989) 264:17627-17630
#title Quail (Coturnix japonica) protamine, full-length cDNA
sequence, and the function and evolution of vertebrate
protamines.
#cross-references MUID:90036816
#accession A34356
#status preliminary
#molecule_type mRNA
#residues 1-57 #label OLI
#cross-references GB:M30275; NID:g213612; PID:g213613
CLASSIFICATION #superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein
#length 57 #molecular_weight 7828 #checksum 5804

Query Match 6.3%; Score 113; DB 2; Length 57;
Best Local Similarity 46.2%; Pred. No. 2.44e-02;
Matches 18; Conservative 7; Mismatches 11; Indels 3; Gaps 3;

Db 16 RRRRRRRRRRRRRGRS-RRSYRSVGRRRRRYGRRRRR 53
||||| : : : : : : : : : : : : : : : : : :
37 5 RRF/RRRRHRRPSHLGQILRRRPWLVP-HPRRY-RWRRK 41

RESULT 9
ENTRY #type complete
TITLE protamine P1 - duckbill platypus
ORGANISM #formal_name Cinithorhynchus anatinus #common_name duckbill
platypus
DATE 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change
29-Jan-1999
ACCESSION# S39425
REFERENCE# S39425
#authors Retief, J.D.; Winkfein, R.J.; Dixon, G.H.
#journal Eur. J. Biochem. (1993) 218:457-461
#title Evolution of the monotremes. The sequences of the protamine
PI genes of platypus and echidna.
#cross-references MUID:94094837
#accession S39425
#status preliminary
#molecule_type DNA
#residues 1-61 #label RET
#cross-references EMBL:Z26849
GENETICS
#introns 47/
CLASSIFICATION #superfamily sperm histone
SUMMARY #length 61 #molecular_weight 7871 #checksum 4490

Query Match 6.3%; Score 113; DB 2; Length 61;
Best Local Similarity 43.8%; Pred. No. 2.44e-02;
Matches 21; Conservative 10; Mismatches 14; Indels 3; Gaps 3;

Db 11 SRSLYRRR-RSR-RGGQRTSRK-LSRRRRGRSRRRRGRSRRSR 55
: ||||| : : : : : : : : : : : : : : : : :
37 4 PRRYRRRRHRRPSHLGQILRRRPWLVP-HPRRYRWRRKNGIENRSLR 51

RESULT 10
ENTRY #type complete
TITLE probable membrane protein YOR053w - yeast (Saccharomyces
cerevisiae)
ALTERNATE_NAMES hypothetical protein O2799
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change
21-Nov-1997
ACCESSION# S66936; S66927
REFERENCE# S66929
#authors Bohn, C.; Bolotin-Fukuhara, M.; Daignan-Fornier, B.; Dang,
D.V.; Valens, M.
#submission submitted to the Protein Sequence Database, July 1996

```

```

KEYWORDS      transmembrane protein
FEATURE
  3-19        #domain transmembrane #status predicted #label TM1\
  107-123     #domain transmembrane #status predicted #label TM2
SUMMARY
    Query Match          6.1%; Score l10; DB 2; Length 126;
    Best Local Similarity 31.7%; Pred. No. 5.99e-02;
    Matches 13; Conservative 14; Mismatches 14; Indels 0; Gaps 0;

Db  49 KRRETRKRRKRKRKRKRKRKRKRKRKRKRKRKRKRKNKDAFY 89
   QY 5 RRRYRHHPRSHLGGQILRRPWLVPVHPHRYWRKXNGIF 45

RESULT 12
ENTRY   C69488             #type complete
TITLE   L50 ribosomal protein L32E (rpL32E) homolog - Archaeoglobobus
        fulgidus
ORGANISM #formal_name Archaeoglobobus fulgidus
DATE     05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
        05-Jun-1998
ACCESSIONS C69488
REFERENCE A93250
AUTHORS Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson,
        K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.;
        Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham,
        D.E.; Kyripides, N.C.; Fleischmann, R.D.; Quackenbush, J.;
        Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.;
        Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftis, B.;
        Petersen, S.; Reich, C.I.; McNeill, L.K.; Badger, J.H.;
        Glödek, A.; Zhou, L.; Overbeek, T.; Gocayne, J.D.; Weidman,
        J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs,
        T.; Artiaich, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.;
        D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.;
        Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese,
        C.R.; Venter, J.C.
#journal Nature (1997) 390:364-370
#title The complete genome sequence of the hyperthermophilic,
        sulfate-reducing archaeon Archaeoglobobus fulgidus.
#cross-references UID:98049343
#accession C69488
#status preliminary; nucleic acid sequence not shown;
        translation not shown
#molecule_type DNA
#residues 1-129 #label KLF
#cross-references GB:AE000971; GB:AE000782; NID:g2689294; PID:g2648647;
        TIGR:AF1908
CLASSIFICATION #superfamily rat ribosomal protein L32
SUMMARY #length 129 #molecular-weight 15151 #checksum 3045

Query Match          6.1%; Score l10; DB 2; Length 129;
Best Local Similarity 31.4%; Pred. No. 5.99e-02;
Matches 16; Conservative 14; Mismatches 18; Indels 3; Gaps 3;

Db  3 HKRLKVRQRQARKPEF-RRYCNWKKLRNKNRSWRPRGLFN-KLRKYG 51
   QY 5 RRRYRHHPRSHLGGQILRRPWLVPVHPHRYR-WRKXNGIFNLSTFG 54

RESULT 13
ENTRY   EDBEIF           #type complete
TITLE   immediate-early protein IE180 - suid herpesvirus 1 (strain
        Indiana-Funkhauser)
ORGANISM #formal_name suid herpesvirus 1
DATE     30-Jun-1990 #sequence_revision 30-Jun-1990 #text_change
        16-Feb-1997
ACCESSIONS S04713
REFERENCE S04713
AUTHORS Cheung, A.K.
#journal Nucleic Acids Res. (1989) 17:4637-4646
#title DNA nucleotide sequence analysis of the immediate-early gene
        of pseudorabies virus.

```



```
#cross-references MWID:8931567
#accession S04713
#molecule_type DNA
#residues 1-1460 #label CHE
CLASSIFICATION #superfamily herpesvirus immediate-early protein IE175
KEYWORDS DNA binding; early protein; transcription regulation
SUMMARY #length 1460 #molecular-weight 152851 #checksum 5209

Query Match          6.1%; Score 110; DB 1; Length 1460;
Best Local Similarity 40.0%; Pred. No. 5,996-02;
Matches      20; Conservative    7; Mismatches 21; Indels   2; Gaps   2;

Ddb      308 RRR-RARRPRPREGRGTRPRRGCA-PLQROPRRRAGEGALRRGRFS 355
        ||| || ||| : ||| : ||| : ||| : ||| : ||| : ||| :
Qy       5 RRYRRRRPRSHLGOILRRRLPWLVHPRHYRWREKNGIENLRSLTFG 54

RESULT 14
ENTRY
TITLE
ORGANISM
DATE
AUTHORS
REFERENCES
```

#authors **SOURCE**
Chauviere, M.; Martinage, A.; Briand, G.; Sautiere, P.;
Chevallier, P.

```

#cross-references #MUID:88055030
#accession S00180
#molecule_type protein
#residues 1-87 ##label CHA
CLASSIFICATION superfamily sperm histone
KEYWORDS DNA binding; nucleus; phosphoprotein; sperm
FEATURE
3,55 #binding_site phosphate (Ser) (covalent) #status
experimental
#length 87 #molecular-weight 11179 #checksum 9188

SUMMARY

Query Match 6.0%; Score 108; DB 2; Length 87;
Best Local Similarity 29.8%; Pred. No. 1.08e-01;
Matches 17; Conservative 14; Mismatches 22; Indels 4; Gaps 4;

Ddb 2 KGYRNRSRPRYRRYGRMRKTRCR-RKGR-RISRPRHTTYRRRY-RKRVHLKRRS 55
: ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 5 RRYRRRRRHRPSHGLQAILR-RPMLVHPHYRRWRKKGIENTRLSRFGYTVKRT 60
: ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      15
ENTRY
TITLE      C7ll159      #type complete
           hypothetical protein PH0472 - Pyrococcus horikoshii
           #formal_name Pyrococcus horikoshii
           #formal_name Pyrococcus horikoshii
DATE      14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change
           14-Aug-1998
C7ll159
REFERENCE
A71000
AUTHORS      Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.;
           Hino, Y.; Yamamoto, S.; Sekine, M.; Baba, S.; Kosugi, H.;
           Hosoyama, A.; Nagai, Y.; Sakai, M.; Ogura, K.; Otsuka, R.;
           Nakazawa, H.; Takamiya, M.; Ohfuku, Y.; Funahashi, T.;
           Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi,
           A.; Aoki, K.; Yoshizawa, T.; Nakamura, Y.; Robb, F.T.;
           Horikoshi, K.; Masuchi, Y.; Shizuya, H.; Kikuchi, H.
           DNA Res. (1998) 5:55-76
#journal
#title      Complete sequence and gene organization of the genome of a.
           hyperthermophilic archaeobacterium, Pyrococcus horikoshii
           OT3.
#cross-references MUID:98344137

```

 W P S R L A

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:14:27 1999; MasPar time 8.36 Seconds
 Tabular output not generated. -787.853 Million cell updates/sec

Title: >US-09-209-961-5
 Description: (1-233) from US09209961.pep
 Perfect Score: 1805
 Sequence: 1 MTPYRRRRRRRRPRSHLG.....RVMTYVQFRFNKDPPLP 233

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 46.711; Variance 82.776; scale 0.564

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.	
1	153	8.5	77	1	PRT2_SEPOF	SPERMATID-SPECIFIC PRO	8.86e-10
2	146	8.1	78	1	PRT1_NORTY	SPERMATID-SPECIFIC PRO	1.43e-08
3	143	7.9	60	1	HSP1_NORTY	SPERMATID-SPECIFIC PRO	4.63e-08
4	139	7.7	63	1	HSP1_DROAU	SPERMATID-SPECIFIC PRO	2.18e-07
5	137	7.6	59	1	HSP1_PHACI	SPERMATID-SPECIFIC PRO	4.71e-07
6	131	7.3	60	1	HSP1_DASVI	SPERMATID-SPECIFIC PRO	4.56e-06
7	131	7.3	61	1	HSP1_SASHA	SPERMATID-SPECIFIC PRO	4.56e-06
8	130	7.2	61	1	HSP1_MACEU	SPERMATID-SPECIFIC PRO	6.63e-06
9	130	7.2	61	1	HSP1_PABBI	SPERMATID-SPECIFIC PRO	6.63e-06
10	130	7.2	61	1	HSP1_TRIVU	SPERMATID-SPECIFIC PRO	6.63e-06
11	130	7.2	62	1	HSP1_MURLO	SPERMATID-SPECIFIC PRO	6.63e-06
12	130	7.2	62	1	HSP1_DASRO	SPERMATID-SPECIFIC PRO	6.63e-06
13	129	7.1	60	1	HSP1_MACGI	SPERMATID-SPECIFIC PRO	9.62e-06
14	129	7.1	61	1	HSP1_MACRG	SPERMATID-SPECIFIC PRO	9.62e-06
15	129	7.1	61	1	HSP1_ANTLA	SPERMATID-SPECIFIC PRO	9.62e-06
16	128	7.1	63	1	HSP1_ANTST	SPERMATID-SPECIFIC PRO	1.39e-05
17	126	7.0	57	1	HSP1_DIDMA	SPERMATID-SPECIFIC PRO	2.91e-05
18	127	7.0	60	1	HSP1_MACAG	SPERMATID-SPECIFIC PRO	2.02e-05
19	126	7.0	68	1	HSP1_PSECU	SPERMATID-SPECIFIC PRO	2.91e-05
20	125	6.9	59	1	HSP1_MACRU	SPERMATID-SPECIFIC PRO	4.20e-05
21	125	6.9	60	1	HSP1_CAEFU	SPERMATID-SPECIFIC PRO	4.20e-05
22	124	6.9	61	1	HSP1_ANTSW	SPERMATID-SPECIFIC PRO	6.06e-05
23	113	6.3	56	1	HSP_COTJA	SPERM HISTONE (PRTAMI	3.01e-03

24 115 6.3 60 1 HSP1_ORNAN SPERMATID-SPECIFIC PRO 3.01e-03
 25 112 6.2 61 1 HSP1_ORNAN SPERMATID-SPECIFIC PRO 4.24e-03
 26 110 5.1 129 1 RL32_ARCFU 50S RIBOSOMAL PROTEIN 8.40e-03
 27 108 5.0 62 1 HSP1_PLATE SPERMATID-SPECIFIC PRO 1.65e-02
 28 109 6.0 87 1 SSS1_SCYCA SPERMATID-SPECIFIC PRO 1.65e-02
 29 107 5.9 238 1 SFR7_HUMAN SPERMATID-SPECIFIC PRO 2.30e-02
 30 106 5.9 284 1 YPV2_METTF HYPOTHETICAL 33.2 KD P 3.21e-02
 31 105 5.8 31 1 HSP2_CUOPA SPERMATID-SPECIFIC PRO 4.47e-02
 32 105 5.8 61 1 HSP2_PLAMS SPERMATID-SPECIFIC PRO 4.47e-02
 33 104 5.8 815 1 CC53_YEAST CELL DIVISION CONTROL 6.21e-02
 34 102 5.7 58 1 HSP3_HORSE SPERMATID-SPECIFIC PRO 1.19e-01
 35 102 5.7 61 1 HSP3_HORSE SPERMATID-SPECIFIC PRO 1.19e-01
 36 102 5.7 62 1 HSP3_HORSE SPERMATID-SPECIFIC PRO 1.19e-01
 37 103 5.7 252 1 VPHE_NPVAC POLYHEDRAL ENVELOPE PR 8.60e-02
 38 103 5.7 449 1 VP1_CAVC1 PROBABLE COAT PROTEIN 1.19e-01
 39 102 5.7 449 1 VP1_CAVC1 PROBABLE COAT PROTEIN 1.19e-01
 40 103 5.7 514 1 VE2_HPV05 REGULATORY PROTEIN E2 8.60e-02
 41 101 5.6 102 1 HSP2_MACMU SPERMATID-SPECIFIC PRO 1.64e-01
 42 101 5.6 104 1 HSP2_CALJA SPERMATID-SPECIFIC PRO 1.64e-01
 43 100 5.5 449 1 VP1_CAVC1 PROBABLE COAT PROTEIN 2.26e-01
 44 100 5.5 449 1 VP1_CAVC1 PROBABLE COAT PROTEIN 2.26e-01
 45 100 5.5 5217 1 HTS1_COCCA HC-TOXIN SYNTHETASE (E 2.26e-01

ALGMENTS

RESULT 1
 ID PR22_SEPOF STANDARD; PRT; 77 AA.
 AC P50002;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE SPERMATID-SPECIFIC PROTEIN T2 [CONTAINS: SPERMATID-SPECIFIC PROTEIN T2]
 XX SEPIA OFFICINALIS (COMMON CUTTLEFISH)
 XX EUPHYOTIA; METAZOA; MOLLUSCA; CEPHALOPODA; COLEOIDA; SEPTOIDA;
 XX SEPIIDAE; SEPIA.
 RN SEQUENCE.
 RP SEQUENCE.
 RX MEDLINE: 91373359.
 RA WOUTERS-TYRROU D., CHARTIER-HARLIN M.-C., MARTIN-PONTHIEU A.,
 EUTILLON C., VAN DORSSLAER A., SAUTIERE P.;
 FT "Cuttlefish spermatid-specific protein T. Molecular characterization
 of two variants T1 and T2, putative precursors of sperm protamine
 variants Sp1 and Sp2."
 RT J. BIOL. CHEM. 266:17388-17395(1991).
 RN [2]
 RP SEQUENCE OF 22-77.
 RX MEDLINE: 91153298.
 RA MARTIN-PONTHIEU A., WOUTERS-TYRROU D., BELAICHE D., SAUTIERE P.,
 SCHINDLER P., VAN DORSSLAER A.;
 FT "Cuttlefish sperm protamines. 1. Amino acid sequences of two distinct
 variants."
 RT EUR. J. BIOCHEM. 195:611-619(1991).
 CC -! FUNCTION: CUTTLEFISH SPERMIOGENESIS IS CHARACTERIZED BY A DOUBLE
 NUCLEAR PROTEIN TRANSITION: HISTONES -> SPERMATID-SPECIFIC
 PROTEINS (T1/T2) -> PROTEIN (SP1/SP2). THE PROTEIN COMPLEX
 SPERMATID-SPECIFIC PROTEIN COMPLEX, STABLE AND INACTIVE COMPLEX.
 CC -! SUBCELLULAR LOCATION: NUCLEAR.
 CC -! TISSUE SPECIFICITY: TESTIS.
 CC -! DEVELOPMENTAL STAGE: SPERMIOGENESIS.
 CC -! PHOSPHORYLATION OCCURS AT DIFFERENT DEGREES. THE
 TRIPHOSPHORYLATED FORM MAY BE PREDOMINANT IN T2. SP2 APPEARS TO BE
 PHOSPHORYLATED IN ELONGATED SPERMATIDS, BUT DEPHOSPHORYLATED IN
 MAJURE SPERM CELLS.
 CC -! SIMILARITY: HIGH, TO S.OFFICINALIS SPERMATID-SPECIFIC PROTEIN T1.
 DR PIR: S14086; S14086.
 DR PIR: B40973; B40973.
 KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
 TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN; PHOSPHORYLATION;
 MULTIGENE FAMILY.
 FT CHAIN 1 77 SPERMATID-SPECIFIC PROTEIN T2.
 FT CHAIN 22 77 SPERMATID-SPECIFIC PROTEIN T2.
 FT DOMAIN 1 1 HYDROPHOBIC.

```

RESULT 3
ID HSPI_NOTTY STANDARD; PRT; 60 AA.
P42143;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SPERM PROTAMINE P1.
GN PRM1.
OS PROTEIN TYPHLOPS (MARSUPIAL MOLE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC NOTORYCTEMORPHIA; NOTORYCTIDAE; NOTORYCTES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE; 95215351.
RT RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L35446; G598338;
CC PROSITE; PS00048; PROTAMINE_P1; 1.
CC ZFAM; PF00260; Protamine_P1; 1.
CC CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
CC TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
CC INIT MET 0 0 BY SIMILARITY.
CC SEQUENCE 60 AA: 8398 MW; 08A902FD CRC32;
CC
CC Query Match. 7.9%; Score 143; DB 1; Length 60;
CC Best Local Similarity 40.8%; Pred. No. 4.63e-08;
CC Matches 20; Conservative 12; Mismatches 16; Indels 1; Gaps 1;
CC
CC 2 RYHRSRSRSRYRRRRRSRYRSORRYRRHRSRGR-RRRGRRGY 49
CC |||||:||||:||||:|:||||:||||:|:|:|
CC 7 RYRRRRPRSHLQGLLRPRPMLVPHRRHRYRRRRKNGIFNRLSRTFGY 55
CC
CC
CC RESULT 4
ID HSPI_DROAU STANDARD; PRT; 63 AA.
P42132;
DT 01-NOV-1995 (REL. 32, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE SPERM PROTAMINE P1.
GN PRM1.
OS DROMICIOPS AUSTRALIS (MONITO DEL MONTE) (DROMICIOPS GLIROIDES).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC MICROBIOTHERIA; MICROBIOTHEIIDAE; DROMICIOPS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPERM;
RX MEDLINE; 95215351.
RT RETIEF J.D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R.J., DIXON G.H.;
RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
RL PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).
CC -!- FUNCTION: PROTAMINES SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
CC SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
CC SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -----

```

P42135; P42133;
01-NOV-1995 (REL. 32, CREATED)
01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
SPERM PROTAMINE P1.
PRM1.
DASYURUS VIVERRINUS (SOUTHEASTERN QUOLL), AND DASYURUS HALLUCATUS.

CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
OC DASYUROMORPHIA; DASYURIDAE; DASYURUS.
RN {}
RN SEQUENCE FROM N.A.
RN RP
RN TISSUE-SPERM;
RX MEDLINE; 95215251.
RX RETIF J. D., KRAJEWSKI C., WESTERMAN M., WINKFEIN R. J., DIXON G. H.;
RT "Molecular phylogeny and evolution of marsupial protamine pl genes.";

PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).
 -!- FUNCTION: PROTEINASE SUBSTITUTES FOR HISTONES IN THE CHROMATIN OF
 SPERM DURING THE HAPLOID PHASE OF SPERMATOGENESIS. THEY COMPACT
 SPERM DNA INTO A HIGHLY CONDENSED, STABLE AND INACTIVE COMPLEX.
 -!- SUBCELLULAR LOCATION: NUCLEAR.
 -!- TISSUE SPECIFICITY: TESTIS.
 This SWISS-PROT entry is covrignt. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL: L3534C; G598103; --
EMBL: L3534A; G598103; --
EMBL: L3534B; G598103; --

```

OR PROSIE; PS000487; PROLAMINE-F1; 1
DR FRAM; P00360; PROLAMINE-P1; 1
KW CHROMOSOMAL PROTEIN; NUCLEOSOME CORE; SPERMATOGENESIS; DNA-BINDING;
KW TESTIS; DNA CONDENSATION; NUCLEAR PROTEIN.
FT INIT.MET 0 0 BY SIMILARITY.
SQ SEQUENC'E 60 AA; 8246 MW; D670FAB3 CRC32;
Query Match . . . 7.3% Score 131; DB 1; Length 60;

```

Best Local Similarity 38.9%; Pred. No. 4.5oe-06;
Matches 21; Conservative 14; Mismatches 15; Indels 4; Gaps 4;

RESULT	7	STANDARD;	PRT;	61 AA.
ID	HEPL_SARHA			
AC	P42151;			
DT	01-NOV-1995	(REL. 32, CREATED)		
DT	01-NOV-1995	(REL. 32, LAST SEQUENCE UPDATE)		
DT	15-DEC-1998	(REL. 37, LAST ANNOTATION UPDATE)		
DE	SPERM PROTAINE PL.			
EN	PRM1.			

SC SARCOPHYLUS HARRISII (TASMANIAN DEVIL), AND DASYVORUS MACULATUS;
CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; METATHERIA;
EN DASYUROMORPHIA; DASYURIDAE; SARCOPHILUS.
EN [1].
RP SEQUENCE FROM N.A.
RC SPECIES-S. HARRISII; TISSUE-SPERM;
RX MEDLINE: 95215351
PA RETIEF, J. D. KRATZWSKI C. WESTERMAN M. WINKEFEN B. J. DIXON G. H.

RT "Molecular phylogeny and evolution of marsupial protamine P1 genes.";
 XL PROC. R. SOC. LOND., B, BIOL. SCI. 259:7-14(1995).
 RN [2]
 RN SEQUENCE FROM N. A.
 RN SPECIES-D. MACULATUS;
 RC KRAJEWSKI C. YOUNG, J., BUCKLEY L., WOOLLEY P. A., WESTERMAN M.;
 RA "Reconstructing the taxonomic radiation of dasyurine marsupials with
 RT RT-PCR using a 25 base pair conserved region of the protamine P1
 RN gene";

AL. CYCLOHOME 7F, 123 FAW, and 260 containing F1 gene trees. ;
J. MANMAG. EVOL. 4:217-236(1997).

[illegible]

Db 2 RYRHSRSRSRYRRRRRRSH-HNRITYRRSRHSR-RRR-GRRGYSRRR 52
OY 7 RYRRR-RHRPRSHLQOILRRRPWLHPHRYRWRKNGIENTLSRTFGYTVKR 59

Search completed: Wed Dec 22 10:14:37 1999
Job time : 10 secs.

[illegible]

```

Title:
Description:
Perfect Score:
Sequence:
Scoring table:

```

Database:

a-geneseq35

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	79	17.8	112 33	W60650	Human C-C chemokine D	2.34e+01
2	77	17.3	91 20	R97665	RANTES	3.41e+01
3	77	17.3	91 13	R70802	RANTES protein.	3.41e+01
4	77	17.3	124 31	R37864	Human cytoplasmic	3.41e+01
5	77	17.3	263 26	R63905	Type I ribosome-inact	3.41e+01
6	77	17.3	263 14	R74179	Type I ribosome-inact	3.41e+01
7	77	17.3	263 7	R37293	Plant type I RIP Momo	3.41e+01
8	76	17.1	310 22	W20383	H. pylori cytoplasmic	4.11e+01
9	76	17.1	326 39	W82311	Human Al adenosine re	4.11e+01
10	76	17.1	326 16	R97655	Human adenosine recep	4.11e+01
11	75	17.1	326 18	R93989	Human ventricule Al ad	4.11e+01
12	75	17.1	326 9	R47386	Human Al adenosine re	4.11e+01
13	76	17.1	326 8	R41524	Human Al adenosine re	4.11e+01
14	76	17.1	326 9	R45744	Human Al adenosine re	4.11e+01
15	76	17.1	326 15	R84192	Human Al adenosine re	4.11e+01
16	76	17.1	446 22	W20936	H. pylori cytoplasmic	4.11e+01

CC can also be used for detection, diagnosis and drug screening.
SQ Sequence 112 AA;

Query Match 17.8%; Score 79; DB 33; Length 112;
Best Local Similarity 37.8%; Pred. No. 2.34e+01;
Matches 17; Conservative 10; Mismatches 14; Indels 4; Gaps 4;

Db 7 fcsllllsllspdaafllpbtacc-tqlyrkplskllrkv 50
QY 11 FC-IFPLTFKSSAPRK-FL-TNWTGCCFATVTRIPLSNKLAV 52

RESULT

ID R97665 standard; Protein; 91 AA.

AC R97665;
DT 14-FEB-1997 (first entry)

DE RANTES.

KW Human; RANTES: Raised on Activation, normal T cell derived and secreted;
KW bone marrow; antagonist; MIP-1alpha; asthma; allergic rhinitis;
KW atopic dermatitis; atheroma; atherosclerosis; rheumatoid arthritis;
KW inflammation; chemotaxis.

OS Homo sapiens.

FH Key Location/Qualifiers

FT peptide 1..23

FT protein /note- "Signal peptide"

FT 24..91

FT /note- "RANTES, claim 3"

FT WO9617935-A2.

PN 13-JUN-1996.

PD 07-DEC-1995; G02861.

PR 16-DEC-1994; GB-024835.

PR 16-JUN-1995; GB-012319.

PA (GLAX) GLAXO GROUP LTD.

PI Proudfoot ABI, Wells TNC;

DR WPI; 96-287180/29;

DR N-PSDB; T30329.

PT Polypeptide functioning as antagonist to RANTES or to MIP-1-alpha
PT used in the treatment of e.g. asthma, allergic rhinitis,
PT atheroma/atherosclerosis or rheumatoid arthritis

PS Example; Fig 1; 33pp; English.

CC This sequence represents full length human RANTES (Raised on Activation,
CC cloned from a human bone marrow lambda gtl cDNA library using the primer
CC sequences given in T30330-33. RANTES was used in the design of
CC peptides which act as antagonists to RANTES or to MIP-1alpha (see
CC also R97666-68). These peptides may be used in the treatment of
CC asthma, allergic rhinitis, atopic dermatitis, atheroma/atherosclerosis
CC or rheumatoid arthritis by inhibiting or reducing inflammation mediated
CC by RANTES or MIP-1alpha. They are also useful in studying RANTES
CC induced chemotaxis, mobilisation of Ca2+ and receptor binding.

SQ Sequence 91 AA;

Query Match 17.3%; Score 77; DB 20; Length 91;

Best Local Similarity 38.5%; Pred. No. 3.41e+01;

Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apasaspyssdtppccfayiar-plp 43

QY 20 SSASPRKFLTNVTCGFATVTRIPLS 45

RESULT

ID R70802 standard; Protein; 91 AA.

AC R70802;

DT 29-AUG-1995 (first entry)

DE RANTES protein.

KW RANTES; T-cell; T-lymphocyte; heparanase; heparin; heparan sulfate;
KW arthritis; restenosis; cancer; wound healing.

OS Homo sapiens.

PN WO9504158-A.

PD 09-FEB-1995.

PF 26-JUL-1994; U08207.

PR 24-JUL-1993; US-099866.

PR 13-OCT-1993; US-136117.

PA (UPJO) UPJOHN CO.

PI Hoogwerf AJ, Ledbetter SR;

DR WPI; 95-082239/11.

DR N-PSDB; Q85372.

PT Screening for cpds. with anti-heparanase activity - by detecting

PT inhibition of heparin or heparan sulphate degradation, cancer.

PT potentially useful for treating arthritis, restenosis, cancer.

PS Claim 13; Page 51; 60pp; English.

CC Purified heparanases, prepared under reducing conditions and

CC activated with transglutaminase, are given in R70786-802. Most

CC are prepared by reverse transcription of mRNA from activated human

CC leukocytes, then cloning of the cDNA into pVL392 baculovirus

CC vector, and expression in Sf9 cells in the presence of reduced

CC glutathione and dithiothreitol.

SQ Sequence 91 AA;

Query Match 17.3%; Score 77; DB 13; Length 91;

Best Local Similarity 38.5%; Pred. No. 3.41e+01;

Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apasaspyssdtppccfayiar-plp 43

QY 20 SSASPRKFLTNVTCGFATVTRIPLS 45

RESULT

ID W37864 standard; Protein; 154 AA.

AC W37864;

DT 10-AUG-1998 (first entry)

DE Human protein comprising secretory signal amino acid sequence 1.

KW Human protein; secretory signal; nutritional source; cytokine;

KW immunity; haematopoiesis; activin; inhibin; tumour; chemotactic;

KW chemokine; thrombolytic; anti-inflammatory; inhibition;

KW fibrosarcoma cell.

OS Homo sapiens.

PN WO9811217-A2.

PD 19-MAR-1998.

PF 12-SEP-1997; J03239.

PR 13-SEP-1996; JP-243060.

PA (PROT-) PROTEGENE INC.

PA (SAGA) SAGAMI CHEM RES CENTRE.

PI Rato S, Kobayashi M, Sekine S, Yamaguchi T;

DR WPI; 98-207380/18.

DR N-PSDB; V29031, V29032.

PT Human proteins with secretory signal sequences - used to treat

PT immune deficiencies, infections, tumours, and haematopoietic

PT disorders, etc.

PS Claim 1; Pages 64-65; 131pp; English.

CC This is the amino acid sequence of a novel human protein comprising

CC a secretory signal isolated from fibrosarcoma cells. Its proteins

CC can be used as nutritional sources or supplements. The proteins may

CC also have cytokine functions, immune modulating functions,

CC haematopoiesis regulating activity, activin/inhibin regulating

CC activity, chemotactic/chemokinetic activity, haemostatic and

CC thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, tumour inhibition activity.

SQ Sequence 154 AA;

Query Match 17.3%; Score 77; DB 31; Length 154;

Best Local Similarity 38.5%; Pred. No. 3.41e+01;

Matches 10; Conservative 9; Mismatches 6; Indels 1; Gaps 1;

Db 19 apasaspyssdtppccfayiar-plp 43

QY 20 SSASPRKFLTNVTCGFATVTRIPLS 45

RESULT

ID R63905 standard; Protein; 263 AA.

AC R63905;

DT 28-JUL-1995 (first entry)

DE Type I ribosome-inactivating protein mormordin II.

PN US5416202-A.
16-MAY-1995. 988430.
PF 09-DEC-1992;
PR 04-NOV-1991; US-787567.
PR 09-DEC-1992; US-988430.
(XOMA) XOMA CORP.
PI Bernhard SP, Better MD, Carroll SF, Lane JA, Lei SP;

```

Db      24  clpfskvydip1lystisdrrfilldltsyvetis 61
        | | : : || : : : | : | : | : | : | :
Qc      4  TLVFSRECFIEPLTKSSASPRKE-LTNVTGCCFATVT 40

```

RESULT	8
ID	W20383 standard; protein; 310 AA.
AC	W20383;
DE	14-JUL-1997 (first entry)
DT	H. pylori cytoplasmic protein, 29844512.aa.
DE	Cytoplasmic; vaccine; prevention; Infection; identification;
KW	binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW	duodenal ulcer disease; chronic gastritis; diagnosis; envelope.
KW	Helicobacter pylori.
OS	Helicobacter pylori.
FH	Key Location/Qualifiers
FT	misc_difference 299
FT	/note= "encoded by RCT"
FN	W09640893-Al.
PD	19-DEC-1996.
PD	06-JUN-1996; U09122.
PR	07-JUN-1995; US-487032.
PR	01-APR-1996; US-630405.
PA	(ASTR) ASTRA AB.
PI	Berglindh OT, Smith D, Mellgaerd BL;
DR	WPI: 97-052306/05.
DR	N-PSDB; T67570.
PT	Helicobacter pylori nucleic acid sequences and related
PT	polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT	infection, and to detect Helicobacter
PS	Claim 61; Page 569; 1481pp; English.
CC	The present sequence is a H. pylori cytoplasmic protein.
CC	The protein may be used in a vaccine to prevent or treat H. pylori
CC	infection or to identify H. pylori polypeptide binding compounds,
CC	useful as potential H. pylori life cycle activators or inhibitors.
CC	The genomic sequence of H. pylori (ATCC 55679) was determined from
CC	overlapping contigs generated by mechanically shearing the bacterial
CC	DNA. The sequences were analysed for ORF of at least 180 nucleotides,
CC	and the predicted coding regions defined by computer evaluation. To
CC	identify likely H. pylori antigens for vaccine development, the amino
CC	acid sequences predicted from various ORF were analysed for significant
CC	homology to other known or exported membrane proteins. Having identified
CC	and determined the sequences of interest, particular regions can be
CC	isolated from H. pylori by PCR amplification for recombinant polypeptide
CC	production, e.g. in E. coli hosts.
SO	Sequence 310 AA;

```

Query Match      17.1%; Score 76; DB 22; Length 310;
Best Local Similarity 22.7%; Pred.No. 4.11e+01;
Matches 10; Conservative 18; Mismatches 15; Indels 1; Gaps 1;

Db 26 lkrslnsanhcvlsagygassaakxf-qeillgvcpisaktkn 68
QV 1 MKCTLVFQSRGFIIPITFKSSASPRFLNVTGCGFATVTRIPL 44
    : : : : : : : : : : : : : : : : : : : : : :
    : : : : : : : : : : : : : : : : : : : : :

```

RESULT	9	
ID	W82311	standard; Protein; 326 AA.
AC	W82311;	
DT	15-MAR-1999	(first entry)
DE	Human A1 adenosine receptor.	
DE	Adenosine receptor A1; human;	cardioprotective; screening; assay;
KW	agonist; antagonist; ischaemia;	angina; myocardial infarction;
KW	therapy.	
OS	Homo sapiens.	
PN	W09850047-A1.	
PD	12-NOV-1998.	
PF	08-MAY-1998;	U090311.
PR	10-OCT-1997;	US-061716.
PR	09-MAY-1997;	US-046030.
PA	(UIPE-) UNIV PENNSYLVANIA,	
PI	Jacobson KA, Liang BT;	
DR	WPI: 99-034683/03.	
DR	N-PSDB: V73320.	
PT	Preventing or reducing	ischaemic injury to heart muscle
PT	activating adenosine A1 and A3	receptors and optionally also
PT	antagonising A2a receptors,	used for drug screening
PS	Example 5; Fig 11A-B;	85pp; English.

```

CC This is the amino acid sequence of human adenosine receptor A1.
CC In a claimed method, ischaemic damage to the heart is prevented or
CC reduced by administering: (i) an agonist that activates both
CC cardiac A1 and A3 adenosine receptors, optionally together with an
CC antagonist that inhibits activity of cardiac A2a adenosine
CC receptors; (ii) a binary conjugate that acts as A3 agonist and A2a
CC antagonist; and (iii) at least one each of A1-specific and
CC A3-specific agonists. The method is used to treat a patient having
CC chronic stable angina, unstable angina or post myocardial infarction
CC angina (claimed). Also claimed are recombinant cardiac myocytes
CC comprising a nucleic acid (see V73320-22) encoding the human A1,
CC A2a or A3 receptor (see W82311-13). The recombinant myocytes are
CC used in a claimed method for determining whether a test compound
CC exerts a cardioprotective effect.
SQ Sequence 326 AA;

Query Match 17.1%; Score 76; DB 39; Length 326;
Best Local Similarity 28.9%; Pred. No. 4.11e+01;
Matches 13...Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 iplykmvtprraaavaiaagcwlsvfvgltpmfgwnnisavera 155
   :|||::|||::|||::|||::|||::|||::|||::|||::
Qy 14 FPLTFKSSASPRFLNTVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 10
ID R87655 standard; Protein: 326 AA.
AC R87655;
DT 19-JUN-1996 (first entry)
DE Human adenosine receptor A1 subtype.
KW in situ hybridisation; adenosine receptor subtype; A3; A1; A2a; A2b;
KW Primer; probe; reverse transcription polymerase chain reaction; RT-PCR;
KW human; eosinophil; antagonist; inhibition; adenylate cyclase;
KW cytokine induced hypersensitivity; allergy; inflammatory disease;
KW asthma; rhinitis; serum sickness; atopic dermatitis; psoriasis; eczema;
KW ulcerative colitis; canine; ventricle; kidney.
```

(MERCER & CO INC.,
PI JACOBSON MA, Johnson RG, Salvatore CA;
DR WPI; 95-360715/47.
DR N-PSDB; T07648.
PT Decreasing intracellular CAMP levels in eosinophil(s) - by
PT contacting cell with antagonist of A3 subtype adenosine receptor.
PS Disclosure; Fig 1; 120pp; English.
CC This sequence represents the human adenosine receptor A1 subtype. RNA
CC was extracted from homogenised human ventricle tissue and probed using
CC synthetic probes based on the canine A1 sequence. The isolated human
CC A1 clone is 92% identical to the canine A1 sequence. The isolated human
CC sequence was amplified and probe fragments were used to screen a human
CC kidney cDNA library. The isolated sequence was used in the
CC construction of the full length A1 cDNA. The full length cDNA may be
CC used in the determination of expression of the A3 adenosine receptor by
CC human eosinophils. This is used in the identification of antagonists of
CC A3 subtype adenosine receptors. These antagonists prevent activation of
CC eosinophils, esp. they prevent A3 inhibition of adenylate cyclase and
CC cytokine induced hypersensitivity. They may be used for treating or
CC preventing allergic or inflammatory diseases, e.g. asthma, rhinitis,
CC serum sickness, atopic dermatitis, psoriasis, eczema, ulcerative colitis
CC etc.
SO Sequence 326 AA;

```

Query Match      ... 17.1% Score 76; DB 16; Length 326;
Best Local Similarity 28.9%; Pred. No. 4,11e+01;
Matches 13; Conservative 16; Indels 3; Gaps 3;

Db 111 ipirykmvncpraaiaagcawlsfvvgtpmfgwnnisavera 155
QY 14 PPTFTKSSASPRFLNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

```

RESULT 11
 ID R3989 standard; Protein; 326 AA.
 AC R3989;
 DT 28-OCT-1996 (first entry)
 DE Human ventricle A1 adenosine receptor.
 KW Human ventricle A1 adenosine receptor; COS7; Xenopus oocyte;
 KW ligand binding; ss.
 OS Homo sapiens.
 PN W09511681-A1.
 PD 04-MAY-1995.
 PF 26-OCT-1994; U12272.
 PR 29-OCT-1993; US-145437.
 PA (MERI) MERCK & CO INC.
 PA (UYVI-) UNIV VIRGINIA PATENTS FOUND.
 PI Doyle MP, Dullin BR, Jacobson MA, Johnson RG, Linden JM;
 DR WPI: 95-178635/23.
 DR N-PSDB: 729929.
 PT Use of xanthine derivs. as adenosine A3 receptor inhibitors - used
 PT to block vasoconstriction, e.g. to treat myocardial ischaemia or
 PT reperfusion injury, inflammation, auto-immune disorders, allergies,
 PT brain arteriole constriction, etc.
 PS Disclosure: Fig 1: 103pp; English.
 CC Human ventricle A1 adenosine receptor has high affinity for
 CC xanthine-agonists with saturated rings in the N6 position of the
 CC adenine ring and xanthine-antagonists with saturated rings in the
 CC C8 position. Recombinant receptor may be produced by subcloning
 CC the cDNA into plasmid pSVL, plasmid pCMV5, or plasmid pREP.
 CC Membranes prepared from transfected COS7 (ATCC CRL 1551) cells
 CC were utilized for the determination of binding affinity,
 CC selectivity and specificity of the human adenosine receptor for
 CC various ligands. A transient expression system in Xenopus oocytes,
 CC established by microinjection of in vitro transcribed mRNA from the
 CC cloned cDNA, also allows measurement of the biological effects upon
 CC activation of the expressed adenosine receptor with ligand binding.
 SQ Sequence. 326 AA;
 Query Match 17.1%; Score 76; DB 18; Length 326;
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;
 Db 111 iplykmvtprraavaiagcwilsvfgltmfgwnnlisavera 155
 QY 14 PPLTFKSSASPRKELTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55
 RESULT 12
 ID R47386 standard; Protein; 326 AA.
 AC R47386;
 DT 29-JUN-1994 (first entry)
 DE Human A1 adenosine receptor.
 KW G protein-coupled receptor; agonist; antagonist; sedative;
 KW epilepsy; bronchoconstriction; hypotensive.
 OS Homo sapiens.
 PN W09325677-A.
 PD 23-DEC-1993.
 PF 11-JUN-1993; AU0277.
 PR 12-JUN-1992; AU-002936.
 PA (GARV-) GARVAN INST MEDICAL RES.
 PI Furlong T, Pierce KD, Seibie L, Shine J, Townsend-Nicholson CA;
 PI Seibie L;
 DR WPI: 94-007535/01.
 DR N-PSDB: Q54640.
 PT DNA sequences encoding the human A1, A2a and A2b adenosine
 PT receptors - for screening for anti-hypertensives, hypnotics,
 PT anti-psychotics, bronchodilators and for recombinant receptor
 PT prodn.
 PS Disclosure: Fig 1: 26pp; English.
 CC The sequence of the human A1 adenosine receptor was isolated from
 CC a human hippocampal cDNA library by screening with specific
 CC consensus oligonucleotides corresponding to the second extracellular
 CC loop (Q54639) and to the third intracellular loop (Q54638). The

CC coding sequence is useful for recombinant production of the A1
 CC receptor to facilitate agonist/antagonist screening. The amino acid
 CC s-sequence R47386 was deduced from the cDNA sequence.
 SQ Sequence. 326 AA;
 Query Match 17.1%; Score 76; DB 9; Length 326;
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;
 Db 111 iplykmvtprraavaiagcwilsvfgltmfgwnnlisavera 155
 QY 14 PPLTFKSSASPRKELTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 13
 ID R41524 standard; Protein; 326 AA.
 AC R41524;
 DT 09-MAR-1994 (first entry)
 DE Human A1 adenosine receptor.
 KW Adenosine receptor; expression system; cardiovascular disorders;
 KW renal disorders; neurological disorders; assay.
 OS Homo sapiens.
 PN GB2264948-A.
 PD 15-SEP-1993.
 PF 05-MAR-1993; 004582.
 PR 13-MAR-1992; US-850701.
 PR 13-MAR-1992; US-850702.
 PR 13-MAR-1992; US-850707.
 PR 15-JAN-1993; US-005945.
 PA (MERI) MERCK & CO INC.
 DR WPI: 93-290442/37.
 DR N-PSDB: Q48414.
 PT Human adenosine receptor proteins - and expression systems,
 PT including cells for use in adenosine receptor binding assay
 PS Claim 5; figure 1; 65pp; English.
 CC Purified adenosine receptor proteins can be used in assays to screen
 CC for therapeutic human adenosine receptor agonists, antagonists or
 CC binding enhancers. These agonists, antagonists and binding
 CC enhancers may be used in the treatment of cardiovascular, renal and
 CC neurological disorders.
 SQ Sequence. 326 AA;
 Query Match 17.1%; Score 76; DB 8; Length 326;
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;
 Db 111 iplykmvtprraavaiagcwilsvfgltmfgwnnlisavera 155
 QY 14 PPLTFKSSASPRKELTNVTGC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 14
 ID R45744 standard; Protein; 326 AA.
 AC R45744;
 DT 25-JUL-1994 (first entry)
 DE Human A1 adenosine receptor.
 KW Adenosine; adenylate cyclase; screening; gene therapy; ischemia;
 KW cystic fibrosis; RFLP.
 OS Homo sapiens.
 PN W09402605-A.
 PD 03-FEB-1994.
 PR 20-JUL-1993; U06782.
 PR 22-JUL-1992; US-918314.
 PA (UYDU-) UNIV DKE.
 PI Olah M, Ren H, Stiles GL;
 DR WPI: 94-048863/06.
 DR N-PSDB: Q55711.
 PT DNA encoding the human A1 adenosine receptor - used for
 PT developing agents for diagnosis or treatment of conditions
 PT involving the receptor such as cystic fibrosis
 PS Disclosure: Fig 2; 48pp; English.
 CC The human A1 adenosine receptor genomic clone was isolated from a
 CC human leukocyte genomic library in EMBL3 using bovine adenosine

CC receptor as probe. The genomic DNA was used as a probe to isolate
 CC human A1 adenosine receptor cDNA clones from a human brain cDNA
 CC library. The DNA encodes a protein which selectively binds
 CC adenosine and inhibits adenylate cyclase. The protein may
 CC be used to develop therapeutic agents to manipulate cardiac function,
 CC protect against ischemia, regulate smooth muscle tone in blood
 CC vessels or for treating cystic fibrosis or seizure activity.
 CC See also R45742-3.
 SQ Sequence 326 AA;

Query Match 17.1%; Score 76; DB 9; Length 326;
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 iplykmvtprraavaiaagcwllsfvvgltptmfgwnnlsavera 155

QY 14 FPLTFKSSASPRKFLTNVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

RESULT 15

ID R84192 standard; Protein; 326 AA.

AC R84192;

DT 25-MAR-1996 (first entry)

DE Human A1 adenosine receptor.

KW Adenosine receptor A1; adenosine receptor agonist;

KN autoimmune disease; therapy.

OS Homo sapiens.

PN GB2289218-A.

PD 15-NOV-1995.

PF 01-MAY-1995; 008844.

PR 06-MAY-1994; US-239473.

PA (MERI) MERCK & CO INC.

PI Jacobson MA;

DR WPI; 95-375695/49.

DR N-PSDB: T00644.

PT Inhibiting prodn. of tumour necrosis factor alpha - with adenosine

PT A2b subtype receptor agonist, useful for treating auto-immune

PT diseases, also method for identifying A2b specific cpds.

PS Disclosure; Fig 1; 96pp; English.

CC Tumor necrosis factor-alpha production, especially in monocytes, is

CC inhibited by contacting the adenosine receptor (especially the A2b

CC subtype) with an adenosine receptor-agonist (adenosine 5'-(N-

CC cyclopropyl)carboxamidoadenosine, 5'-(N-ethyl)carboxamidoadenosine,

CC (R)-N6-phenyl-2-propyladenosine or cyclohexyladenosine. The

CC agonists may be used in the therapy of autoimmune diseases,

CC specifically rheumatoid arthritis, rheumatoid spondylitis,

CC inflammatory bowel disease (ulcerative colitis and Crohns disease),

CC intestinal pathology associated with graft vs. host disease, organ

CC transplant reactions, septic shock, fever and myalgia due

CC to infection and cachexia associated with chronic infections,

CC malignancy and AIDS, pulmonary diseases e.g. sarcoidosis,

CC silicosis, chronic pulmonary inflammatory disease, and adult

CC respiratory distress syndrome.

SQ Sequence 326 AA;

Query Match 17.1%; Score 76; DB 15; Length 326;
 Best Local Similarity 28.9%; Pred. No. 4.11e+01;
 Matches 13; Conservative 13; Mismatches 16; Indels 3; Gaps 3;

Db 111 iplykmvtprraavaiaagcwllsfvvgltptmfgwnnlsavera 155

QY 14 FPLTFKSSASPRKFLTNVTC-CFA-TVTRIPLSN-KVLTAVDRS 55

Search completed: Wed Dec 22 10:19:28 1999
 Job time : 18 secs.

Result No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	101	24.3	222	1	YORA_PYRWO	HYPOTHETICAL 24.7 KD P	1.52e+02
2	93	22.4	233	1	ATPT_CYACA	ATP SYNTHASE A CHAIN P	1.88e-01
3	91	21.9	1574	1	MYOSIN_2	MYOSIN-2 ISOFORM.	3.46e-01
4	89	21.4	544	1	MANB_MYCPI	PHOSPHOMANNOMUTASE (EC	6.32e-01
5	87	20.9	1254	1	MDR3_CAEEL	MULTIDRUG RESISTANCE P	1.15e+00
6	87	20.9	1411	1	YK63_CAEEL	HYPOTHETICAL 157.6 KD	1.15e+00
7	85	20.4	1226	1	YNAL_HORVU	PROBABLE INTEGRIN ALPHA	2.06e+00
8	84	20.2	438	1	AMYL_HORVU	ALPHA-AMYLASE TYPE A I	2.75e+00
9	83	20.0	456	1	PUR8_HAEIN	ADENYLOSUCCINATE LYASE	3.66e+00
10	83	20.0	830	1	PKH1_YEAST	MTK1 PROTEIN.	3.66e+00
11	82	19.7	203	1	YG12_BACVU	HYPOTHETICAL 22.8 KD P	4.86e+00
12	82	19.7	352	1	YORM_TTVI1	HYPOTHETICAL 38.6 KD P	4.86e+00
13	82	19.7	382	1	RFC_SHIFL	O-ANTIGEN POLYMERASE.	4.86e+00
14	81	19.5	317	1	MSHR_HUMAN	MELANOCYTE STIMULATING	6.45e+00
15	81	19.5	459	1	ME31_DROME	PUTATIVE ATP-DEPENDENT	6.45e+00
16	81	19.5	715	1	VGL_SPVIR	CAPSID PROTEIN.	6.45e+00
17	81	19.5	1068	1	YCF0_MARPO	HYPOTHETICAL 127 KD PR	6.45e+00
18	80	19.2	124	1	RFBI_SHIFL	RFBI PROTEIN.	8.53e+00
19	80	19.2	214	1	SC22_YEAST	PROTEIN TRANSPORT PROT	8.53e+00
20	80	19.2	240	1	Y03M_MYCTU	HYPOTHETICAL 25.7 KD P	8.53e+00
21	80	19.2	1095	1	NKC2_MOUSE	BUMETANIDE-SENSITIVE S	8.53e+00
22	80	19.2	1095	1	NKC2_RAT	BUMETANIDE-SENSITIVE S	8.53e+00
23	79	19.0	246	1	CDM_HUMAN	CDM PROTEIN (6C6-AG TU	1.12e+01

RX MEDLINE: 94033298
 RA KOSTRZEWA M., ZETSCHKE K.;
 RT "Organization of plastid-encoded ATPase genes and flanking regions
 including homologues of infB and tsf in the thermophilic red alga
 Galdieria sulphuraria.";
 RL PLANT MOL. BIOL. 23:67-76(1993).
 CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
 CC DIRECT ROLE IN THE TRANSLLOCATION OF PROTONS ACROSS THE MEMBRANE.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CHLOROPLAST
 CC THYLAKOID MEMBRANE.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X67814; G429173;
 DR PIR: S39515; S39515.
 DR PROSITE: PS00449; ATPASE.A; 1.
 DR PFAM: PF00119; ATP-synt_A; 1.
 DR MENDEL: 537; CVACA:atp1.1.
 KW HYDROGEN ION TRANSPORT; CF(0); CHLOROPLAST; TRANSMEMBRANE; SIGNAL.
 FT SIGNAL 1 ? POTENTIAL.
 FT CHAIN ? 233 ATP SYNTHASE A CHAIN.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 121 141 POTENTIAL.
 FT TRANSMEM 190 210 POTENTIAL.
 FT TRANSMEM 211 231 POTENTIAL.
 SQ SEQUENCE 233 AA; 26066 MW; 8407838F CRC32;

Query Match 22.4%; Score 93; DB 1; Length 233;
 Best Local Similarity 33.3%; Pred. No. 1.88e-01;
 Matches 11; Conservative 14; Mismatches 7; Indels 1; Gaps 1;

Db 11 HFFVTLGVFLK-HAQVFIVSFFVILILFSLV 42
 QY 6 HLGFKRGVFKFSELYIHGYTDIWLWVTFV 38

RESULT 3
 ID MYS2-YEAST STANDARD; PRT; 1574 AA.
 AC P19524;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE MYOSIN-2 ISOFORM.
 GN MYO2 OR CD266 OR YOR326W OR O6167.
 OS SACCCHAROMYCES CEREVISIAE (BAKER'S YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTA; HEMIASCOCETES; SACCCHAROMYCETES;
 OC SACCCHAROMYCETACEAE; SACCCHAROMYCES.
 [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRF88;
 RX MEDLINE: 91201404.
 RA JOHNSTON G.C., PRENDERGAST J.A., SINGER R.A.;
 RT "The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin
 for vectorial transport of vesicles.";
 RL J. CELL BIOL. 113:539-551(1991).
 RN [2]

RP SEQUENCE OF 1-748 FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE: 97051589.
 RA PEARSON B.M., HERNANDO Y., PAYNE J., WOLF S.S., KALOGEROPOULOS A.,
 RA SCHWEIZER M.;
 RT "Sequencing of a 35.71 kb DNA segment on the right arm of yeast

RT chromosome XV reveals regions of similarity to chromosomes I and...
 RT XIII.;
 RL YEAST 12:1021-1031(1996).
 RN [3]
 RP SEQUENCE OF 677-1574 FROM N.A.
 RX MEDLINE: 97051586.
 RA PARLE-MCDERMOTT A.G., HAND N.J., GOULDING S.G., WOLFE K.H.;
 RT "Sequence of 29 kb around the PDR10 locus on the right arm of
 RT Saccharomyces cerevisiae chromosome XV: similarity to part of
 RT chromosome I.";
 RL YEAST 12:999-1004(1996).
 CC -1- FUNCTION: ESSENTIAL MYOSIN FOR VECTORIAL TRANSPORT OF VESICLES.
 CC PLAYS A DETERMINANT ROLE IN THE SPATIAL REGULATION OF BUD
 CC FORMATION.
 CC -1- SUBUNIT: MAY SELF-ASSOCIATE.
 CC -1- CA(2+) MAY PLAY A ROLE IN MYO2 ACTIVITY.
 CC -1- SIMILARITY: BELONGS TO CLASS-5 MYOSINS. STRONGEST, TO YEAST MYO4.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: M35532; G172022;
 DR EMBL: X90565; G940860;
 DR EMBL: Z49821; E218684;
 DR EMBL: Z75234; E252153;
 DR EMBL: Z75235; E332815;
 DR PIR: A38454; A38454.
 DR SGD: L0001223; MYO2.
 DR PFAM: PF00063; myosin_head; 1.
 DR HSP: P08799; LMND.
 KW MYOSIN; COILED COIL; ATP-BINDING; ACTIN-BINDING;
 KW HEPTAD REPEAT PATTERN; ALKYLATION; MULTIGENE FAMILY; PHOSPHORYLATION;
 KW CALMODULIN-BINDING.
 FT DOMAIN 1 781 GLOBULAR HEAD-LIKE DOMAIN.
 FT DOMAIN 782 925 NECK.
 FT DOMAIN 926 981 COILED COIL.
 FT DOMAIN 982 1009 HINGE.
 FT DOMAIN 1010 1086 COILED COIL.
 FT DOMAIN 1087 1574 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
 FT DOMAIN 443 523 ACTIN-BINDING (BY SIMILARITY).
 FT DOMAIN 750 925 ARG/LYS-RICH (BASIC).
 FT DOMAIN 791 805 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 814 828 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 839 853 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 862 876 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 887 902 CALMODULIN-BINDING (BY SIMILARITY).
 FT DOMAIN 910 924 CALMODULIN-BINDING (BY SIMILARITY).
 FT NP_BIND 164 171 ATP (BY SIMILARITY).
 FT MOD_RES 692 692 ALKYLATION (BY SIMILARITY).
 FT MOD_RES 702 702 ALKYLATION (BY SIMILARITY).
 FT MOD_RES 1097 1097 PHOSPHORYLATION (BY CDC2) (POTENTIAL).
 SQ SEQUENCE 1574 AA; 180680 MW; AB9D381A CRC32;

* Query Match 21.9%; Score 91; DB 1; Length 1574;
 Best Local Similarity 34.4%; Pred. No. 3.46e-01;
 Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

Db 90 IKQVSYSLNIVTYSIV-LIATNPFDRVDQLY 120
 QY 14 FKIKFSELYIHGYTDIWLWVTFVFSAEAY 45

RESULT 4
 ID MANB_MYCPI STANDARD; PRT; 544 AA.
 AC P47723;
 DT 01-FEB-1996 (REL. 33, CREATED)
 DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)


```

CC      (ABC TRANSPORTERS). BELONGS TO THE MDR SUBFAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to licensed@isb-sib.ch).
CC
CC      EMBL: X65055; G6811;
CC      DR      S27338; S27338;
CC      FR      PROSITE; P560211; ABC_TRANSPORTER; 2.
CC      DR      PFAM; PF000005; ABC_tran; 2.
CC      FR      PFAM; PF00664; ABC_membrane; 2.
CC      KW      ATP-BINDING; GLYCOPROTEIN; TRANSMEMBRANE; TRANSPORT; DUPLICATION;
CC      MULTIGENE FAMILY.
CC      DOMAIN      1      46      CYTOPLASMIC (POTENTIAL).
CC      FT      TRANSFM      47      67      POTENTIAL.
CC      FT      TRANSFM      104      124      POTENTIAL.
CC      FT      TRANSFM      173      193      POTENTIAL.
CC      FT      TRANSFM      200      220      POTENTIAL.
CC      FT      TRANSFM      281      301      POTENTIAL.
CC      FT      TRANSFM      310      330      POTENTIAL.
CC      FT      TRANSFM      331      693      CYTOPLASMIC (POTENTIAL).
CC      FT      DOMAIN      694      714      POTENTIAL.
CC      FT      TRANSFM      738      758      POTENTIAL.
CC      FT      TRANSFM      814      834      POTENTIAL.
CC      FT      TRANSFM      835      855      POTENTIAL.
CC      FT      TRANSFM      918      938      POTENTIAL.
CC      FT      TRANSFM      955      975      POTENTIAL.
CC      FT      DOMAIN      976      1254      CYTOPLASMIC (POTENTIAL).
CC      FT      NP_BIND      411      418      ATP (POTENTIAL).
CC      FT      NP_BIND      1050      1057      ATP (POTENTIAL).
CC      FT      SEQUENCE      1254 aa: 138807 MW; A9AD92D2 CRC32;
CC
CC      Query Match      20.9%; Score 87; DB 1; Length 1254;
CC      Best Local Similarity 36.4%;
CC      Matches      8; Conservative      6; Mismatches      8; Gaps      0;

```

```

00      .98 LVEFLRGFTWPAFSVIGULF / 19
       :| | :| | :| |
QY      1 MVFIHLGFRKGVFKIKFSELY 22

RESULT        6
ID   YK63_CAEEL    STANDARD;          PRT;   1411 AA.
AC   P34342;
DT   01-FEB-1994 (REL. 28, CREATED)
DT   01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT   01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE   HYPOTHETICAL 137.5 KD PROTEIN C29E4.3 IN CHROMOSOME III.
GN   C29E4.3.
OS   CAENORHABDITIS ELEGANS.
OC   EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
CC   RHABDITINA; RHABDITOIDEA; PELODERINAE; CAENORHABDITIS.
EN   [1]
RP   SEQUENCE FROM N.A.
PC   STRAIN-BR-STOL N2;
AX   MEDLINE; 94150718.
QA   WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
BA   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA   CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FRASER A.,
FA   FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
JA   JOHNSTON-L., JONES M., KERSHAW J., KIESTEN J., LAISTER N.,
LA   LATREIJE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
KA   PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEN R.,
RA   SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
RA   SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
RA   WATERSCYN R., WAYSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
RA   WOHLDRUP P.;
PT   RA
RT   "2.2 Mb cf contiguous nucleotide sequence from chromosome III of C.
     elegans.";
```

RL NATURE 368:32-38(1994).
 CC -!- SIMILARITY: TO MAMMALIAN GALACTOCEREBROSIDASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L23651; G388581;
 CC PIR: S44770; S44770; CE00085.
 CC WORMPEP: F54F2.1; CE00194.
 CC PFAM: PF00560; LRR; 5.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 1411 AA; 157562 MW; FD6C6D14 CRC32;

Query Match 20.98; Score 87; DB 1; Length 1411;
 Best Local Similarity 34.28; Pred. No. 1.15e+00;
 Matches 13; Conservative 11; Mismatches 10; Indels 4; Gaps 4;

Db 1093 RFINEDYDIIYENLV-IFSAFO-STOKYITSHVKGK 1128
 QY 17 KFSELY-IHGTDIVLVVTFVRSAAIVV-HISRG 52

RESULT 7
 ID YMAL CAEEL STANDARD; PRT; 1226 AA.
 AC P34416;
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE PROBABLE INTEGRIN ALPHA CHAIN F54F2.1 PRECURSOR.
 GN F54F2.1
 OS CAENORHABDITIS ELEGANS.
 OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
 OC RHABDITIA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE; 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU Z., DURIN R., FAYELLO A., FRASER A.,
 RA FULTON L., GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M.,
 RA JOHNSTON L., JONES M., KERSHAW J., KIRSTEN J., LAISTER N.,
 RA LATREILLE P., LIGHTNING J., LLOYD C., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SIMS M., SMALDON N., SMITH A., SMITH M., SONNHAMMER E., STADEN R.,
 RA SULSTON J., THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K.,
 RA WATERSON R., WATSON A., WEINSTOCK L., WILKINSON-SPROAT J.,
 RA WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL NATURE 368:32-38(1994).
 CC -!- FUNCTION: POSSIBLE ROLE IN CELL-CELL INTERACTIONS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INTEGRIN ALPHA CHAIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: L23645; G388605;
 CC PIR: S44824; S44824.
 CC WORMPEP: F54F2.1; CE00194.
 CC PROSITE: PS00242; INTEGRIN_ALPHA; 1.
 CC PFAM: PF00357; integrin_A; 2.

DR HSEPL P11215; I88X.
 KW HYPOTHETICAL PROTEIN; INTEGRIN; CELL ADHESION; GLYCOPROTEIN;
 KW TRANSMEMBRANE; SIGNAL; EXTRACELLULAR MATRIX; CYTOSKELETON.
 FT SIGNAL 1 ? 1226
 FT CHAIN ? 1154
 FT DOMAIN ? 1154
 FT TRANSEM 1155 1177
 FT DOMAIN 1178 1226
 FT CARBOHYD 108 108
 FT CARBOHYD 228 228
 FT CARBOHYD 290 290
 FT CARBOHYD 608 608
 FT CARBOHYD 679 679
 FT CARBOHYD 775 775
 FT CARBOHYD 819 819
 SQ SEQUENCE 1226 AA; 135939 MW; A9AC1B00 CRC32;

Query Match 20.48; Score 85; DB 1; Length 1226;
 Best Local Similarity 27.38; Pred. No. 2.06e+00;
 Matches 9; Conservative 12; Mismatches 12; Indels 0; Gaps 0;

Db 384 GDLNLDGYNDVIGVAPYAGKNGKQAVYVHGSK 416
 QY 19 SELYHGTYDVLVWTFVRSAAIVVHISR 51

RESULT 8
 ID AMV1 HORVU STANDARD; PRT; 438 AA.
 AC P00693;
 DT 21-JUL-1986 (REL. 01, CREATED)
 DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE ALPHA-AMYLASE TYPE A ISOZYME PRECURSOR (EC 3.2.1.1) (1,4-ALPHA-D-
 DE GLUCAN GLUCANOHYDROLASE) (AMV1) (LOW PI ALPHA-AMYLASE).
 GN AMV1.1
 OS HORDEUM VULGARE (BARLEY).
 OC EUKARYOTA; VIRIDIPANTAE; STREPTOPHYTA; EMBRYOPHYTA; TRACHEOPHYTA;
 OC EUPHYLOTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILLOPSIDA; POALES;
 OC POACEAE; HORDEUM.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. HIMALAYA;
 RX MEDLINE; 83238423.
 RA ROGERS J.C., MILLMAN C.;
 RT "Isolation and sequence analysis of a barley alpha-amylase cDNA
 RT clone.";
 RL J. BIOL. CHEM. 258:8169-8174(1983).
 CC -!- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-ALPHA-GLUCOSIDIC
 CC LINKAGES IN OLIGOSACCHARIDES AND POLYSACCHARIDES.
 CC -!- COFACTOR: BINDS A CALCIUM ION REQUIRED FOR ITS ACTIVITY.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- DEVELOPMENTAL STAGE: PRODUCTION OF ALPHA-AMYLASE IS HORMONALLY
 CC REGULATED. GERMINATING EMBRYOS PRODUCE THE HORMONE GIBBERELLIC
 CC ACID, WHICH WITHIN 10 HOURS STIMULATES THE ALEURONE CELLS COVERING
 CC THE ENDOSPERM OF THE SEED TO PRODUCE ALPHA-AMYLASE. THE ENZYME
 CC THEN DEGRADATES THE STARCH WITHIN THE ENDOSPERM FOR USE BY THE
 CC DEVELOPING PLANT EMBRYO.
 CC -!- THERE ARE AT LEAST 4 TYPES OF ALPHA-AMYLASE IN BARLEY.
 CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
 CC KNOWN AS 'THE ALPHA-AMYLASE FAMILY'.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: J01236; G166987;
 CC PIR: A00846; ALBH.
 CC PFAM: PF00128; alpha-amylase; 1.
 CC MENDEL: 8617; HORVU; AMV1.1.

[illegible]

OG PLASMID PG12.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 OC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H1.1;
 RX MEDLINE; 89098342.
 RA WAHLON J., SEURINCK J.;
 RT "Complete nucleotide sequence of pGI2, a Bacillus thuringiensis
 RL plasmid containing Tn4430.";
 RL NUCLEIC ACIDS RES. 16:11827-11827(1988).
 CC -!- FUNCTION: POSSIBLY INVOLVED IN PG12 REPLICATION MECHANISM.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X13481; G40320; -
 DR PIR; S02048; S02048.
 KW HYPOTHETICAL PROTEIN; PLASMID.
 SQ SEQUENCE 203 AA; 22826 MW; CA914B9A CRC32;
 Query Match 19.7%; Score 82; DB 1; Length 203;
 Best Local Similarity 22.5%; Pred. No. 4.86e+00;
 Matches 9; Conservative 15; Mismatches 15; Indels 1; Gaps 1;
 DB 110 WGGFTLKAGDIFITNATSSAGIVGHAIRA-NGDNYILHMP 148
 QY 11 WGVFKIKFSELYINGYTDIVLVVVFVFERSAEYVHHIS 50
 RESULT 12
 ID YORM.TTV1 STANDARD; PRT; 352 AA.
 AC P19237;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DE HYPOTHETICAL 38.6 KD PROTEIN.
 OS THERMOPROTEUS TENAX VIRUS 1 (STRAIN KRAL) (TTV1).
 OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; LIPOTHRIXVIRIDAE;
 OC LIPOTHRIXVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA NEUMANN H.;
 RL SUBMITTED (MAR-1989) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X14855; G62178; -
 KW HYPOTHETICAL PROTEIN
 SQ SEQUENCE 352 AA; 38674 MW; 51D0560D CRC32;
 Query Match 19.7%; Score 82; DB 1; Length 352;
 Best Local Similarity 23.8%; Pred. No. 4.86e+00;
 Matches 10; Conservative 16; Mismatches 14; Indels 2; Gaps 2;
 DB 125 LFIVQ-GYPWMYNNPVVFAIPGYTGLFWMGIDNVVMMRT 165
 QY 2 VFIIHLGFKGKVFKEISELYIH-GYTDIVLVVVFVFERSA 42
 RESULT 13
 ID RFC_SHIFL STANDARD; PRT; 382 AA.

P37784;
 DT 01-OCT-1994 (REL. 30, CREATED)
 DT 01-OCT-1994 (REL. 30, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE O-ANTIGEN POLYMERASE.
 GN RFC.
 OS SHIGELLA FLEXNERI
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 OC SHIGELLA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SEROTYPE 2A;
 RX MEDLINE; 94131953.
 RA MORONA R., MAVRIS M., FALLARINO A., MANNING P.A.;
 RT "Characterization of the rfc region of Shigella flexneri.";
 RL J. BACTERIOL. 176:733-747(1994).
 CC -!- FUNCTION: MAY LINK THE O-ANTIGEN TETRASACCHARIDE UNITS INTO LONG
 CC CHAINS. GIVING RISE TO TYPICAL SMOOTH LPS.
 CC -!- PATHWAY: LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE
 CC (POTENTIAL).
 CC -!- SIMILARITY: TO OTHER O-ANTIGEN POLYMERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X71970; G454905; -
 KW LIPOPOLYSACCHARIDE BIOSYNTHESIS; TRANSMEMBRANE; INNER MEMBRANE.
 FT TRANSMEM 11 31 POTENTIAL.
 FT TRANSMEM 36 56 POTENTIAL.
 FT TRANSMEM 61 81 POTENTIAL.
 FT TRANSMEM 87 107 POTENTIAL.
 FT TRANSMEM 126 146 POTENTIAL.
 FT TRANSMEM 166 186 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 228 248 POTENTIAL.
 FT TRANSMEM 303 323 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 SQ SEQUENCE 382 AA; 43723 MW; 91B8562A CRC32;
 Query Match 19.7%; Score 82; DB 1; Length 382;
 Best Local Similarity 27.8%; Pred. No. 4.86e+00;
 Matches 10; Conservative 12; Mismatches 11; Indels 3; Gaps 3;
 DB 257 VRFNDLYFY-YKN-VDLATF-LFGKFGFSFLDLRLR 289
 QY 16 IKFSELYINGYTDIVLVVVFVFERSAEYVHHIS 51
 RESULT 14
 ID MSHR.HUMAN STANDARD; PRT; 317 AA.
 AC Q01726;
 DT 01-JUL-1993 (REL. 26, CREATED)
 DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE MELANOCYTE STIMULATING HORMONE RECEPTOR (MSH-R) (MELANOTROPIN
 DE RECEPTOR) (MELANOCORTIN-1 RECEPTOR) (MCL-R).
 GN MCLR OR MSHR
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SKIN.
 RX MEDLINE; 92390715.
 RA MOUNTJOY K.G., ROBBINS L.S., MORTRUD M., CONE R.D.;
 RT "The cloning of a family of genes that encode the melanocortin

Tue Jan 4 11:52:24 2000

US-09-209-961-21.rsp

Page 8

Search completed: Wed Dec 22 10:29:37 1999
Job time : 8 secs.

```

db      113 IISVAFRFSWVYSPFWAILTFNTTGTGVDGLVLFLVLSKMPSDAIVV 161
              : : | : | : | | | | | | : : | |
CQ      1 MVEIHLGFKMGVKIKFSELYHGYTDIVV-LVVFTVER-SAEAYVV 47

2
RESULT          #type complete
ENTRY           H+-transporting ATP synthase (EC 3.6.1.34) chain a - red alga
TITLE           (Cyanidium caldarium) chloroplast
ORGANISM        #formal_name Chloroplast Cyanidium caldarium
DATE            07-Oct-1994 #sequence_revision 10-Nov-1995 #text_change
                14-Nov-1997
ACCESSIONS      S39515
REFERENCE        Koztzeva, M.; Zetsche, K.
AUTHOR'S         Plant Mol. Biol. (1993) 23:67-76
JOURNAL          Organization of plastid-encoded ATPase genes and flanking
TITLE            regions including homologues of infB and tsf in the
                thermophilic red alga Galdieria sulphuraria.
#cross-references MUID:94033298
#accession       S39515
#status          nucleic acid sequence not shown; translation not shown
#molecule type DNA

```

```
ACCESSIONS A38454; S58339; S62058; S67232; S67233; S71966; S72009
REFERENCE- A38454
#authors Johnston, G.C.; Prendergast, J.A.; Singer, R.A.
#journal J. Cell Biol. (1991) 113:539-551
#title The Saccharomyces cerevisiae MYO2 gene encodes an essential myosin for vectorial transport of vesicles.
#cross-references MUID:91201404
#accession A38454
##molecule_type DNA
##residues 1-1574 #label JOH
##cross-references GB:M35532; NID:g172021; PID:g172022
REFERENCE S58318
#authors Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
#submission submitted to the EMBL Data Library, August 1995
#accession S58339
##molecule_type DNA
##residues 1-748 #label PEA
##cross-references EMBL:X90565; NID:g940836; PID:g940860
REFERENCE S62058
#authors Parle, A.G.; Hand, N.J.; Goulding, S.G.; Wolfe, K.H.
#submission submitted to the EMBL Data Library, June 1995
#description Sequence of 29 kilobases around the PDR10 locus on the right arm of Saccharomyces cerevisiae chromosome XV: similarity to part of chromosome I.
#accession S62058
##molecule_type DNA
##residues 677-1574 #label PAR
##cross-references EMBL:Z49821; NID:g1163062; PID:e218684; PID:g1163063
REFERENCE S67213
#authors Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67232
##molecule_type DNA
##residues 1-748 #label PEW
##cross-references EMBL:Z75234; MIPS:YOR326w
##experimental_source strain S288c
REFERENCE S67233
#authors Goulding, S.E.; Hand, N.J.; Parle-McDermott, A.G.; Wolfe, K.H.
#submission submitted to the Protein Sequence Database, July 1996
#accession S67233
##molecule_type DNA
##residues 677-1574 #label GOU
##cross-references EMBL:Z75234; MIPS:YOR326w
##experimental_source strain S288c
REFERENCE S71966
#authors Parle-McDermott, A.G.; Hand, N.J.; Goulding, S.E.; Wolfe, K.H.
#journal Yeast (1996) 12:999-1004
#title Sequence of 29 kb around the PDR10 locus on the right arm of Saccharomyces cerevisiae chromosome XV: similarity to part of chromosome I.
#cross-references MUID:97051586
#accession S71966
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 677-1574 #label PAW
##cross-references EMBL:Z49821; NID:g1163062; PID:e218684; PID:g1163063
##note the nucleotide sequence was submitted to the EMBL Data Library, June 1995
REFERENCE S71986
#authors Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.
#journal Yeast (1996) 12:1021-1031
#title Sequencing of a 35.71 kb DNA segment on the right arm of yeast chromosome XV reveals regions of similarity to chromosomes I and XIII.
#accession S72009
##status nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-748 #label PEF
```



```

#cross-references EMBL:X90565; NID:g940836; PID:g940860
#note the nucleotide sequence was submitted to the EMBL Data
Library, August 1995
COMMENT The neck domain comprises six approximately 23-residue tandem
repeats; this domain may be responsible for calmodulin binding.
GENETICS
#gene SGD:MYO2
#map_position 15R
#cross-references SGD:S0005853; MIPS:YOR326W
FUNCTION
#description involved in vacuole inheritance; involved in polarized growth
and secretion; plays a role in vesicle transport along
actin cables to the bud site
CLASSIFICATION #superfamily myosin MYO2; myosin motor domain homology
KEYWORDS actin binding; ATP; coiled coil; cytoskeleton; hydrolase;
P-loop; phosphoprotein; tandem repeat
FEATURE
73-769 #domain myosin motor domain homology #label MMOT\
164-171 #region nucleotide-binding motif A (P-loop)\
552-673 #region actin binding #status predicted\
782-926 #domain neck #status predicted #label NEC\
942-1086 #domain coiled coil #status predicted #label COI\
1087-1574 #domain carboxyl-terminal #status predicted #label CTD\
170 #binding_site ATP (lys) #status predicted\
692,702 #active_site Cys #status predicted\
1097,1452 #binding_site phosphate (Thr) (covalent) #status
predicted
SUMMARY #length 1574 #molecular-weight 180679 #checksum 13
Query Match 21.9%; Score 91; DB 1; Length 1574;
Best Local Similarity 34.4%; Pred. No. 1.38e+00;
Matches 11; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
Db 90 IKRYSQNLNITYGIV-LIATNPDRVDOLY 120
Qy 14 KRIKFSLYIHGTDIVLVVFTVFERSAEY 45

RESULT 5
ENTRY #type complete
TITLE phosphomannomutase (EC 5.4.2.8) - Mycoplasma pirum (strain
BER) (SGC3)
ORGANISM #formal_name Mycoplasma pirum
DATE 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change
10-Sep-1997
ACCESSIONS E53312
REFERENCE A53312
#authors Tham, T.N.; Ferris, S.; Kovacic, R.; Montagnier, L.;
Blanchard, A.
#journal J. Bacteriol. (1993) 175:5281-5285
#title Identification of Mycoplasma pirum genes involved in the
salvage pathways for nucleosides.
#cross-references MUID:93352438
#accession E53312
#status preliminary
#molecule_type DNA
#residues 1-544 #label THA
#cross-references GB:L13289; NID:g401781; PID:g401786
GENETICS
#genetic_code SGC3
KEYWORDS intramolecular transferase; isomerase
SUMMARY #length 544 #molecular-weight 62108 #checksum 7566
Query Match 21.4%; Score 89; DB 2; Length 544;
Best Local Similarity 36.2%; Pred. No. 2.36e+00;
Matches 17; Conservative 11; Mismatches 15; Indels 4; Gaps 3;
Db 362 VYRGTGFKWGDINKIKDSEFFVGFEEAVGALNSTI-NRDKDAY 407
Qy 2 VFIHLGFW-G-VFKIKFSELYIHGTDIVLVVFTVFERSAEY 45

RESULT 6
ENTRY #type complete
TITLE ABC-type transport protein slr0977 - Synecocystis sp.
(strain PCC 6803)
ALTERNATE_NAMES #formal_name Synecocystis sp.
ORGANISM PCC 6803
#variety
DATE 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change
21-Aug-1998
ACCESSIONS S74745
REFERENCE S74322
#authors Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.;
Nakamura, Y.; Miyajima, N.; Hirosewa, M.; Sugita, M.;
Sasamoto, S.; Kimura, T.; Hosouchi, T.; Matsuno, A.;
Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo,
S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.;
Yasuda, M.; Tabata, S.
#journal DNA Res. (1996) 3:109-136
#title Sequence analysis of the genome of the unicellular
cyanobacterium Synecocystis sp. PCC6803. II. Sequence
determination of the entire genome and assignment of
potential protein-coding regions.
#cross-references MUID:97061201
#accession S74745
#status nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues 1-276 #label KAN
#cross-references EMBL:D90901; GB:AB001339; NID:g1651897; PID:d1017629;
PID:g1651970
#note the nucleotide sequence was submitted to the EMBL Data
Library, June 1996
KEYWORDS transport protein
SUMMARY #length 276 #molecular-weight 31542 #checksum 9527
Query Match 20.9%; Score 87; DB 2; Length 276;
Best Local Similarity 28.9%; Pred. No. 4.01e+00;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;
Db 224 MVSVID-GFRWAILGGE-STIFLPGFL-LSLLLVIIIF 258
Qy 1 MYPILHGFKWGVFKIKFSELYIHGTDIVLVVFTVF 38

RESULT 7
ENTRY #type fragment
TITLE cellulase (EC 3.2.1.4) 88K precursor - Prevotella ruminicola
(fragment)
ALTERNATE_NAMES endo-1,4-beta-glucanase
ORGANISM #formal_name Prevotella ruminicola
DATE 28-May-1992 #sequence_revision 28-May-1992 #text_change
20-Mar-1998
ACCESSIONS A41375; B41375
REFERENCE A41375
#authors Matsushita, O.; Russell, J.B.; Wilson, D.B.
#journal J. Bacteriol. (1991) 173:6919-6926
#title A Bacteroides ruminicola 1,4-beta-D-endoglucanase is encoded
in two reading frames.
#cross-references MUID:92041578
#accession A41375
#status preliminary
#molecule_type DNA
#residues 1-535 #label MAT
#accession B41375
#status preliminary
#molecule_type DNA
#residues 150-535 #label MA2
#cross-references GB:S61753; NID:g238044; PID:g238046
FUNCTION
#description hydrolysis of 1,4-beta-D-glucosidic linkages in
beta-D-glucans such as cellulose and lichenin; can
hydrolyze such linkages in beta-D-glucans that also contain
1,3-linkages
#pathway cellulose degradation
#keywords glycosidase; hydrolase; polysaccharide degradation

```

```

#introns
22/1; 115/1; 188/3; 309/2; 363/1; 596/3; 848/1; 882/3; 920/3
945/3; 1058/3; 1119/1; 1161/3; 1181/1; 1290/2; 1323/1
#length 1411 #molecular-weight 157562 #checksum 9164
SUMMARY

```

```
Matches 10; Conservative 13; Mismatches 18; Indels 0; Gaps 0;

Dt 2 GYMKMFTNLAKESDFDRTMYTSHTQMEGNYELKVAEG 42
      |||..|||:::|:::|:::|:::|:::|:::|
Q/ 12 GVFKIKSELXHGVTIDVILVVVFERSAEAYVVISRG 52

RESULT 14
ENTRY TITLE ALB. #type complete
ORGANISM #formal_name Hordeum vulgare #common_name barley
DATE 03-Aug-1984 #sequence_revision 03-Aug-1984 #text_change
ACCESSIONS A00846; S65602
REFERENCE Rogers, J.C.; Milliman, C.
#authors J. Biol. Chem. (1983) 258:8169-8174
#journal Isolation and sequence analysis of a barley alpha-amylase
#title .cDNA clone.
#cross-references MUID:83238423
#accession A00846
#molecule_type mRNA
#residues 1-438 #label ROG
#cross-references GB:J01236; NID:g166987
#experimental_source cv. Himalaya
REFERENCE Sjuge, N.; Rodenburg, K.W.; Guo, X.J.; Chaix, J.C.; Svensson, B.
#authors FEBS Lett. (1995) 363:299-303
#journal Isozyme hybrids within the protruding third loop domain of
#title the barley alpha-amylase (beta/alpha)(8)-barrel.
#cross-references MUID:95255567
#accession S65602
#molecule_type protein
#residues 25-29 #label JUG
COMMENT Production of this enzyme in barley is hormonally regulated.
Germinating barley embryos produce gibberellic acid, which
stimulates cells covering the endosperm to produce alpha-amylase.
FUNCTION
description catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic
bonds
pathway glycogen/starch degradation
CLASSIFICATION superfamily wheat alpha-amylase; alpha-amylase core homology
KEYWORDS calcium binding; glycosidase; hydrolase; monomer;
polysaccharide degradation
FEATURE
1-24 #domain signal sequence #status predicted #label SIG\
25-438 #product alpha-amylase #status predicted #label MAT\
171-318 #domain alpha-amylase core homology #label AMY\
204,229,315 #active_site Asp, Glu, Asp #status experimental
SUMMARY #length 438 #molecular-weight 47796 #checksum 9025

Query Match : 20.2%; Score 84; DB 1; Length 438;
Best Local Similarity 28.6%; Pred.No. 8.74e+00;
Matches 14; Conservative 12; Mismatches 20; Indels 3; Gaps 3;

Db 346 IFYDF-FNWG-FKOIAALVAIRKNGITATSALKILMHGDYVAEI 392
      | | | | | | | | | | | | | | | | | | | | | |
Q/ 2 VFIIHLGPKWGVKIFKSELY-IHGYTDIVLVVVFERSAEAYVVIH 49

RESULT 15
ENTRY TITLE S51344 #type complete
ORGANISM probable membrane protein YLR341w - yeast (Saccharomyces cerevisiae)
ALTERNATE_NAMES hypothetical protein L8300.7
ORGANISM #formal_name Saccharomyces cerevisiae
DATE 23-Feb-1995 #sequence_revision 12-May-1995 #text_change
ACCESSIONS S51344
REFERENCE Du, Z.
#authors
```

```
#submission submitted to the EMBL Data Library, December 1994
#description The sequence of S. cerevisiae cosmid 8300.
#accession S51344
#molecule_type DNA
#residues 1-477 #label DUZ
#cross-references EMBL:U19028; NID:g609380; PID:g609386; MIPS:YLR341w
GENETICS
#map_position 12R
KEYWORDS transmembrane protein
FEATURE
209-225 #domain transmembrane #status predicted #label TM1\
339-355 #domain transmembrane #status predicted #label TM2
SUMMARY #length 477 #molecular-weight 55138 #checksum 131

Query Match 20.2%; Score 84; DB 2; Length 477;
Best Local Similarity 34.3%; Pred. No. 8.74e+00;
Matches 12; Conservative 13; Mismatches 7; Indels 3; Gaps 3;

Db 351 FILIVKFMYSVNFLEKFSDEVFL-SFIEILIKIVF 384
QY 3 FIHLGFKWGV-FKIRFS-ELYINGYTDIVLVVF 35
```

Search completed: Wed Dec 22 10:29:13 1999
Job time : 10 secs.

Query Match 19.5%; Score 81; DB 1; Length 317;
Best Local Similarity 39.3%; Pred. No. 1.99e+01;
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195
QY 11 WGVFKIKFSELYIHGYTDIVVLVVFVTF 38

RESULT 2
ID US-08-842-045-2 STANDARD; PRT; 317 AA.

XX AC xxxxxx

Sequence 2, Application US/08842045

Sequence 2, Application US/08842045
Patent No. 5817787

GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka
TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/842,045

CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Smith, Deann F.
REGISTRATION NUMBER: 36683
REFERENCE/DOCKET NUMBER: 2115-000853DVE
TELEPHONE: (810)641-1600
TELEFAX: (810)641-0270

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 317 AA; 34691 MW; 523647 CN;

Query Match 19.5%; Score 81; DB 2; Length 317;
Best Local Similarity 39.3%; Pred. No. 1.99e+01;
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195
QY 11 WGVFKIKFSELYIHGYTDIVVLVVFVTF 38

RESULT 3
ID US-07-866-979-6 STANDARD; PRT; 317 AA.

XX AC xxxxxx

Sequence 6, Application US/07866979

XX Sequence 6, Application US/07866979
CC Patent No. 5532347
CC GENERAL INFORMATION:
CC APPLICANT: Cone, Roger D
CC APPLICANT: Mountjoy, Kathleen G
CC TITLE OF INVENTION: Melanocyte Stimulating Hormone Receptor
CC NUMBER OF SEQUENCES: 6
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Allegretti & Witcoff, Ltd.
CC STREET: 10 South Wacker Drive, Suite 3000
CC CITY: Chicago
CC STATE: Illinois
CC COUNTRY: USA
CC ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: 19920410
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: No. 5532347nan, Kevin E
REGISTRATION NUMBER: 35,303
REFERENCE/DOCKET NUMBER: 92,154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-715-1000
TELEFAX: 312-715-1234
TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 317 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE 317 AA; 34560 MW; 525101 CN;

Query Match 19.5%; Score 81; DB 1; Length 317;
Best Local Similarity 39.3%; Pred. No. 1.99e+01;
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 W-VASVVFSTLFIAVDHVAVLCLVVF 195
QY 11 WGVFKIKFSELYIHGYTDIVVLVVFVTF 38

RESULT 4
ID US-08-671-525B-2 STANDARD; PRT; 317 AA.

XX AC xxxxxx

Sequence 2, Application US/08671525B

Sequence 2, Application US/08671525B
Patent No. 5703220
GENERAL INFORMATION:
APPLICANT: Yamada, Tadataka

TITLE OF INVENTION: Genes Encoding Melanocortin Receptors
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
STREET: P.O. Box 828
CITY: Bloomfield Hills
STATE: MI
COUNTRY: US
ZIP: 48303

```

CC --SEQUENCE CHARACTERISTICS:
CC LENGTH: 317 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 317 AA; 34691 MW; 523647 CN;
SQ
Query Match          19.5%; Score 81; DB 2; Length 317;
Best Local Similarity 39.3%; Pred. No. 1.99e+01;
Matches              11; Conservative 5; Mismatches 11; Indels 1;

Db      169 W-VASVVFSTLFIAVDHVAVLCLVWF 195
       |.|.:|||.|.:||.:||.:||
Cc      11 WGVFKIFSELGYTHGTYDIWLWLVTF 38

PEPULT        6
Cp      US-08-780-749A-4    STANDARD; PRT; 317 AA.
Xx      xxxxxx
Dt      XX
Dt      XX
DE      Sequence 4, Application US/08780749A
Xx      XX
Cc      Sequence 4, Application US/08780749A
Cc      Patent No. 5932779
Cc      GENERAL INFORMATION:
Cc      APPLICANT: Lee, Frank
Cc      APPLICANT: Huszar, Dennis
Cc      TITLE OF INVENTION: SCREENING METHODS FOR COMPOUNDS
Cc      TITLE OF INVENTION: USEFUL IN THE REGULATION OF BODY WEIGHT
Cc      NUMBER OF SEQUENCES: 10
Cc      CORRESPONDENCE ADDRESS:
Cc      ADDRESSEE: Pennie & Edmonds LLP
Cc      STREET: 1155 Avenue of the Americas
Cc      CITY: New York
Cc      STATE: New York
Cc      COUNTRY: USA
Cc      ZIP: 10036/2711
Cc      COMPUTER READABLE FORM:
Cc      MEDIUM TYPE: Diskette
Cc      COMPUTER: IBM Compatible
Cc      OPERATING SYSTEM: DOS
Cc      SOFTWARE: FastSeq Version 2.0
Cc      CURRENT APPLICATION DATA:
Cc      APPLICATION NUMBER: US/08/780,749A
Cc      FILING DATE: 08-JAN-1997
Cc      CLASSIFICATION: 800
Cc      ATTORNEY/AGENT INFORMATION:
Cc      NAME: Laura A. Coruzzi
Cc      REGISTRATION NUMBER: 30,742
Cc      REFERENCE/DOCKET NUMBER: 7853-064
Cc      TELECOMMUNICATION INFORMATION:
Cc      TELEPHONE: (212) 790-9090
Cc      TELEFAX: (212) 869-8864/9741
Cc      TELEX: 66141 PENNIE
Cc      INFORMATION FOR SEQ ID NO: 4:
Cc      SEQUENCE CHARACTERISTICS:
Cc      LENGTH: 317 amino acids
Cc      TYPE: amino acid
Cc      STRANDEDNESS:
Cc      TOPOLOGY: unknown
Cc      MOLECULE TYPE: peptide
Cc      SEQUENCE 317 AA; 34691 MW; 523647 CN;
SQ
Query Match          19.5%; Score 81; DB 2; Length 317;
Best Local Similarity 39.3%; Pred. No. 1.99e+01;
Matches              11; Conservative 5; Mismatches 11; Indels 1;

b      169 W-VASVVFSTLFIAVDHVAVLCLVWF 195

```

88

TITLE OF INVENTION: NOVEL HUMAN G-PROTEIN COUPLED RECEPTOR

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: Smithline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696,770

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Han, William T

REGISTRATION NUMBER: 34,344

REFERENCE/DOCKET NUMBER: TAK50001-2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-5219

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 404 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

SEQUENCE 404 AA; 45323 MW; 927330 CN;

Query Match 19.0%; Score 79; DB 2; Length 404;

Best Local Similarity 34.4%; Pred. No. 2.79e+01;

Matches 11; Conservative 9; Mismatches 9; Indels 3; Gaps 3;

Db 121 WGSFSCRTHYFVNMYSIFFLVCLSV-DR 151

QY 11 WGVFKIFSE-LY-IGHYTDIVLVVFTVFER 40

RESULT 10

ID US-08-118-906-14 STANDARD; PRT; 400 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 14, Application US/08118906

Sequence 14, Application US/08118906

Patent No. 5484590

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

APPLICANT: Bierhuizen, Marti F.A.

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a

TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/118,906

FILING DATE: 09-SEP-1993

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Campbell, Cathryn A.

REGISTRATION NUMBER: 51,815

REFERENCE/DOCKET NUMBER: P-LJ 9526

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 535-9001

TELEFAX: (619) 535-8949

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 400 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

TELECOMMUNICATION INFORMATION:

TELEPHONE: 400 AA; 45854 MW; 855321 CN;

Query Match 18.8%; Score 78; DB 1; Length 400;

Best Local Similarity 25.0%; Pred. No. 3.31e+01;

Matches 6; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

Db 3 LFTISVSSVIFIVFVSFVNFQDP 31

QY 21 LYTHGYTDIVLVVFTVFERSAEA 44

RESULT 11

ID US-08-485-196-14 STANDARD; PRT; 400 AA.

XX xxxxxx

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Sequence 14, Application US/08486196

Sequence 14, Application US/08486196

Patent No. 5751420

GENERAL INFORMATION:

APPLICANT: Fukuda, Minoru

APPLICANT: Bierhuizen, Marti F.A.

TITLE OF INVENTION: Expression of the Developmental I

TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a

TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700

CITY: San Diego

STATE: California

COUNTRY: USA

ZIP: 92122

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/486,196

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/118,906

FILING DATE: 09-SEP-1993

CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 9526
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 400 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SQ SEQUENCE 400 AA; 45854 MW; 855321 CN;

Query Match 18.8%; Score 78; DB 1; Length 400;
Best Local Similarity 25.0%; Pred. No. 3.3le+01;
Matches 6; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

Dd 8 LFIISVSVIIIVFSVFNGDP 31
QY 21 LYIHGYTDIIVLVVTVERSAEA 44
| : | : : : |||| : : :
| : | : : : |||| : : :

RESULT 12
ID US-08-488-135-14 STANDARD; PRT; 400 AA.
XX
XX
AC xxxxxx
DT
DE
DE
DE
XX
XX
Sequence 14, Application US/08488135
Sequence 14, Application US/08488135
Patent No. 5766910
GENERAL INFORMATION:
CC APPLICANT: Fukuda, Minoru
CC TITLE OF INVENTION: Expression of the Developmental I
CC TITLE OF INVENTION: Antigen By a Cloned Human cDNA Encoding a Member of a
CC TITLE OF INVENTION: Beta-1,6-N-Acetylglucosaminyltransferase Gene Family
CC NUMBER OF SEQUENCES: 14
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Campbell and Flores
CC STREET: 4370 La Jolla Village Drive, Suite 700
CC CITY: San Diego
CC STATE: California
CC COUNTRY: USA
CC ZIP: 92122
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patentin Release #1.0, Version #1.25
CC CURRENT APPLICATION NUMBER: US/08/488.135
CC FILING DATE:
CC CLASSIFICATION: 424
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/118,906
CC FILING DATE: 09-SEP-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Campbell, Cathryn A.
CC REGISTRATION NUMBER: 31,815
CC REFERENCE/DOCKET NUMBER: P-LJ 9526
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (619) 535-9001
CC TELEFAX: (619) 535-8949
CC INFORMATION FOR SEQ ID NO: 14:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 400 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear

RESULT 14
ID US-08-426-819A-36 STANDARD; PRT; 584 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX Sequence 36, Application US/08426819A
XX
XX Sequence 36, Application US/08426819A
CC Patent No. 5723318
CC GENERAL INFORMATION:
CC APPLICANT: Yamaguchi, No. 5723318omi
CC APPLICANT: Kojima, Tetsuo
CC APPLICANT: Oh-Eda, Masayoshi
CC APPLICANT: Hattori, Kunihiro
CC TITLE OF INVENTION: Genes Coding for Megakaryocyte
CC TITLE OF INVENTION: Potentiator
CC NUMBER OF SEQUENCES: 37
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch
CC STREET: P.O. Box 747
CC CITY: Falls Church
CC STATE: Virginia
CC COUNTRY: USA
CC ZIP: 22040-0747
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/426.819A
CC FILING DATE: 21-APR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Murphy Jr., Gerald M.
CC REGISTRATION NUMBER: 28,977
CC REFERENCE/DOCKET NUMBER: 230-107P
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-205-8000
CC TELEFAX: 703-205-8050
CC INFORMATION FOR SEQ ID NO: 36:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 584 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: not relevant
CC TOPOLOGY: not relevant
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: internal
SQ SEQUENCE 584 AA; 64098 MW; 1744236 CN;

Query Match 18.5%; Score 77; DB 1; Length 584;
Best Local Similarity 47.4%; Pred. No. 3.92e+01;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 313 VLKHKLDLYPQGYPESVI 331
Qy 13 VFKIKFSELYIHGYTDIV 31

RESULT 15
ID US-08-426-819A-35 STANDARD; PRT; 622 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX Sequence 35, Application US/08426819A
XX
XX Sequence 35, Application US/08426819A
CC Patent No. 5723318

CC GENERAL INFORMATION:
CC APPLICANT: Yamaguchi, No. 5723318omi
CC APPLICANT: Kojima, Tetsuo
CC APPLICANT: Oh-Eda, Masayoshi
CC APPLICANT: Hattori, Kunihiro
CC TITLE OF INVENTION: Genes Coding for Megakaryocyte
CC TITLE OF INVENTION: Potentiator
CC NUMBER OF SEQUENCES: 37
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch
CC STREET: P.O. Box 747
CC CITY: Falls Church
CC STATE: Virginia
CC COUNTRY: USA
CC ZIP: 22040-0747
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/426.819A
CC FILING DATE: 21-APR-1995
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Murphy Jr., Gerald M.
CC REGISTRATION NUMBER: 28,977
CC REFERENCE/DOCKET NUMBER: 230-107P
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-205-8000
CC TELEFAX: 703-205-8050
CC INFORMATION FOR SEQ ID NO: 35:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 622 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 622 AA; 68037 MW; 1973385 CN;

Query Match 18.5%; Score 77; DB 1; Length 622;
Best Local Similarity 47.4%; Pred. No. 3.92e+01;
Matches 9; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 351 VLKHKLDLYPQGYPESVI 369
Qy 13 VFKIKFSELYIHGYTDIV 31

Search completed: Wed Dec 22 10:30:33 1999
Job time : 7 secs.

W P S R L H

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:14:55 1999; MasPar time 37.19 Seconds
649.311 Million cell updates/sec
Tabular output not generated.

Title: >US-09-209-961-3
Description: (1-314) from US09209961.pep
Perfect Score: 2358
Sequence: 1 MPSKNGRSGPQPHKRWFT.....GGQFVTLSPCPPEFYINY 314

Scoring table: PAM 150
Gap 11

Searched: 541047 seqs, 76911114 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-pending

1: P9 2: U60 3: U7 4: U80 5: U81 6: U82 7: U83 8: U84A 9: U84B
10: U85 11: U86 12: U87 13: U88 14: U89 15: U90 16: U91 17: U92
18: U93 19: NEWP 20: NEWU6 21: NEWU8 22: NEWU9

Statistics: Mean 36.699; Variance 134.433; scale 0.273

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2358	100.0	314	22	US-09-209- Sequence 3, Applicatio	8.53e-255
2	1988	84.3	312	22	US-09-209- Sequence 4, Applicatio	1.35e-211
3	1988	84.3	312	17	US-09-267- Sequence 21, Applicati	1.35e-211
4	1988	84.3	312	1	PCT-US99-0 Sequence 21, Applicati	1.35e-211
5	648	27.5	220	1	PCT-US99-0 Sequence 8, Applicatio	4.19e-57
6	648	27.5	220	17	US-09-267- Sequence 8, Applicatio	4.19e-57
7	197	8.4	286	9	US-08-484- Sequence 16, Applicati	3.05e-08
8	193	8.2	285	9	US-08-484- Sequence 17, Applicati	7.57e-08
9	139	5.9	286	9	US-08-484- Sequence 18, Applicati	1.07e-02
10	98	4.2	561	16	US-09-107- Sequence 5323, Applic	3.77e+01
11	97	4.1	340	18	US-09-328- Sequence 5334, Applic	4.54e+01
12	96	4.1	411	16	US-09-107- Sequence 4942, Applic	5.45e+01
13	97	4.1	611	1	PCT-US99-0 Sequence 4, Applicatio	4.54e+01
14	97	4.1	511	11	US-08-600- Sequence 4, Applicatio	4.54e+01
15	94	4.0	120	20	US-60-162- Sequence 574, Applicat	7.86e+01
16	92	3.9	284	12	US-08-751- Sequence 8, Applicatio	1.13e+02
17	91	3.9	409	20	US-60-164- Sequence 3663, Applic	1.13e+02
18	92	3.9	479	20	US-60-164- Sequence 66, Applicati	1.13e+02
19	92	3.9	971	2	US-60-096- Sequence 19531, Applic	1.13e+02
20	92	3.9	971	17	US-09-248- Sequence 19531, Applic	1.13e+02
21	92	3.9	2227	9	US-08-475- Sequence 2, Applicatio	1.13e+02

22	52	3.9	2227	7	US-08-397- Sequence 2, Applicatio	1.13e+02
23	52	3.9	2227	21	US-08-475- Sequence 6, Applicatio	1.13e+02
24	52	3.9	2227	7	US-08-397- Sequence 2, Applicatio	1.13e+02
25	52	3.9	2227	7	US-08-397- Sequence 4, Applicatio	1.13e+02
26	52	3.9	2227	21	US-08-475- Sequence 2, Applicatio	1.13e+02
27	85	3.8	332	16	US-09-134- Sequence 4723, Applic	1.93e+02
28	30	3.8	353	20	US-60-164- Sequence 12463, Applic	1.62e+02
29	30	3.8	400	20	US-60-164- Sequence 12370, Applic	1.62e+02
30	30	3.8	420	15	US-09-079- Sequence 223, Applicat	1.62e+02
31	89	3.8	456	10	US-08-566- Sequence 2, Applicatio	1.93e+02
32	89	3.8	456	10	US-08-566- Sequence 2, Applicatio	1.93e+02
33	89	3.8	456	10	US-08-566- Sequence 2, Applicatio	1.93e+02
34	89	3.8	456	10	US-08-566- Sequence 2, Applicatio	1.93e+02
35	30	3.8	510	16	US-09-134- Sequence 5141, Applic	1.93e+02
36	30	3.8	808	20	US-60-167- Sequence 18734, Applic	1.93e+02
37	90	3.8	1056	15	US-09-079- Sequence 217, Applicat	1.62e+02
38	90	3.8	1130	10	US-08-519- Sequence 6, Applicatio	1.62e+02
39	90	3.8	1437	20	US-60-164- Sequence 4032, Applic	1.93e+02
40	90	3.8	1463	12	US-08-747- Sequence 3, Applicatio	1.62e+02
41	90	3.8	2227	21	US-08-475- Sequence 4, Applicatio	1.62e+02
42	90	3.8	2227	9	US-08-475- Sequence 4, Applicatio	1.62e+02
43	90	3.8	4536	15	US-09-079- Sequence 1, Applicatio	1.62e+02
44	90	3.8	4563	16	US-09-108- Sequence 1, Applicatio	1.62e+02
45	90	3.8	553	16	US-09-108- Sequence 1, Applicatio	1.62e+02

ALIGNMENTS

RESULT 1
ID US-09-209-961-3 STANDARD; PRT: 314 AA.
XX
AC .xxxxxx

DE Sequence 3, Application US/09209961

XX Sequence 3, Application US/09209961

CC GENERAL INFORMATION:

CC APPLICANT: WANG, LI

CC APPLICANT: BABIOK, LORNE A.

CC APPLICANT: POTTER, ANDREW A.

CC APPLICANT: WILLSON, PHILIP

CC TITLE OF INVENTION: POSTWEANING MULTISYSTEM WASTING SYNDROME VIRUS FROM

CC TITLE OF INVENTION: PIGS

CC FILE REFERENCE: 9000-0040

CC CURRENT APPLICATION NUMBER: US/09/209,961

CC CURRENT FILING DATE: 1998-12-10

CC EARLIER APPLICATION NUMBER: 60/069,233

CC EARLIER FILING DATE: 1997-12-11

CC EARLIER APPLICATION NUMBER: 60/069,750

CC EARLIER FILING DATE: 1997-12-16

CC NUMBER OF SEQ ID NOS: 24

CC SOFTWARE: PatentIn Ver. 2.0

CC SEQ ID NO: 3

CC LENGTH: 314

CC TYPE: FAT

CC ORIGIN: Porcine Circovirus Type II

CC SEQUENCE 314 AA: 35793 MW: 558374 CN;

Query Match: 100.0%; Score 2358; DB 22; Length 314;
Best Loca.: Similarity 100.0%; Pctd. No. 8.53e-255;
Matches 314; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MPSKNGRSGPQPHKRWFTLNPNSEDERKKIRELPSLFDYFIVGEEGNEGRTPHLOQ 60

1 MPSKNGRSGPQPHKRWFTLNPNSEDERKKIRELPSLFDYFIVGEEGNEGRTPHLOQ 60

Db 61 FANFVKOTFNKVKYLGARCHIEKAKGTDQONKEYCSKGNLLIECGAPRSGQSDLS 120

61 FANFVKOTFNKVKYLGARCHIEKAKGTDQONKEYCSKGNLLIECGAPRSGQSDLS 120

Db 121 TAVSILLESGLITVAEQHPVTKNFKRGIAELLKLLSGKMKRDWTKNVHFIIVGPPCCGK 180

```

121 TAVSTLESILVTVAQHPVTFVKNFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGK 180
181 SKWAANFANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVKTK 240
181 SKWAANFANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVKTK 240
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGQFVT 300
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGQFVT 300
301 LSPPCPEFFYEINY 314
301 LSPPCPEFFYEINY 314

RESULT 2
ID US-09-209-961-4 STANDARD; PRT; 312 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 4, Application US/09209961
Sequence 4, Application US/09209961
GENERAL INFORMATION:
APPLICANT: WANG, LI
APPLICANT: BABIUK, LORNE A.
APPLICANT: POTTER, ANDREW A.
APPLICANT: WILLSON, PHILIP
TITLE OF INVENTION: POSTWEANING MULTISYSTEM WASTING SYNDROME VIRUS FROM
FILE REFERENCE: 9000-0040
CURRENT APPLICATION NUMBER: US/09/209,961
CURRENT FILING DATE: 1998-12-10
EARLIER APPLICATION NUMBER: 60/069,233
EARLIER FILING DATE: 1997-12-11
EARLIER APPLICATION NUMBER: 60/069,750
EARLIER FILING DATE: 1997-12-16
NUMBER OF SEQ ID NOS: 24
SOFTWARE: Patent Ver. 2.0
SEQ ID NO 4
LENGTH: 312
TYPE: PRT
ORGANISM: Porcine Circovirus Type I
SEQUENCE 312 AA; 35704 MW; 527440 CN;

Query Match 84.3%; Score 1988; DB 22; Length 312;
Best Local Similarity 84.1%; Pred. No. 1.35e-211;
Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPKRWVFTLNPNSEEEKNIRELPISLDFYFVCGEEGLEEGRTPHLOGFANFAK 64
QY 8 RSGPQPKRWVFTLNPNSEEEKNIRELPISLDFYFVCGEEGLEEGRTPHLOGFANFAK 67
Db 65 QTNKVKWYLGARCHIEKAKGTDOONKEYCSKEGHILIECGAPRQGRSDLSSTAVSTLL 124
QY 68 QTNKVKWYLGARCHIEKAKGTDOONKEYCSKEGHILIECGAPRQGRSDLSSTAVSTLL 127
Db 125 ETGSLVTVAEQFPVTVYVNRFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGKSWARNF 184
QY 128 ESGILVTVAEQHPVTFVKNFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGKSWARNF 187
Db 185 AEPRDTYKPSRNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVETKGTVPFL 244
QY 188 ANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVETKGTVPFL 247
Db 245 ARSLITSNQAPQEWYSSSTAVPAVEALYRRTITLQFWKTAGEQSTVEPEGRFEAVDPPCA 304
QY 248 ARSLITSNQTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTE-EGGQFVTLSPPCP 306
Db 305 LFPYKINY 312
QY 307 EFPYEINY 314

RESULT 4
ID PCT-US99-05485-21 STANDARD; PRT; 312 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 21, Application PC/TUS9905485
Sequence 21, Application PC/TUS9905485
GENERAL INFORMATION:
```

```

121 TAVSTLESILVTVAQHPVTFVKNFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGK 180
181 SKWAANFANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVKTK 240
181 SKWAANFANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVKTK 240
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGQFVT 300
241 GGTVPFLARSILITSNOTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTEGGQFVT 300
301 LSPPCPEFFYEINY 314
301 LSPPCPEFFYEINY 314

RESULT 3
ID US-09-267-177-21 STANDARD; PRT; 312 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 21, Application US/09267177
Sequence 21, Application US/09267177
GENERAL INFORMATION:
APPLICANT: Poet, Steven E.
APPLICANT: Ritchie, Branson W.
APPLICANT: Niagro, Frank D.
APPLICANT: Lukert, Phil D.
TITLE OF INVENTION: Vaccines against Circovirus Infections
FILE REFERENCE: 21099.0057
CURRENT APPLICATION NUMBER: US/09/267,177
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: 60/077,890
EARLIER FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 21
LENGTH: 312
TYPE: PRT
ORGANISM: porcine circovirus
SEQUENCE 312 AA; 35704 MW; 527440 CN;

Query Match 84.3%; Score 1988; DB 17; Length 312;
Best Local Similarity 84.1%; Pred. No. 1.35e-211;
Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPKRWVFTLNPNSEEEKNIRELPISLDFYFVCGEEGLEEGRTPHLOGFANFAK 64
QY 8 RSGPQPKRWVFTLNPNSEEEKNIRELPISLDFYFVCGEEGLEEGRTPHLOGFANFAK 67
Db 65 QTNKVKWYLGARCHIEKAKGTDOONKEYCSKEGHILIECGAPRQGRSDLSSTAVSTLL 124
QY 68 QTNKVKWYLGARCHIEKAKGTDOONKEYCSKEGHILIECGAPRQGRSDLSSTAVSTLL 127
Db 125 ETGSLVTVAEQFPVTVYVNRFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGKSWARNF 184
QY 128 ESGILVTVAEQHPVTFVKNFRGLAELLKVSGRMOKRDKNTVHFIIVGPPCGKSWARNF 187
Db 185 AEPRDTYKPSRNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVETKGTVPFL 244
QY 188 ANPETTYWKPKNKWDYHGEKVVWVDDFYGLPMDLLRLCDRYPLTVETKGTVPFL 247
Db 245 ARSLITSNQAPQEWYSSSTAVPAVEALYRRTITLQFWKTAGEQSTVEPEGRFEAVDPPCA 304
QY 248 ARSLITSNQTPLEWYSSSTAVPAVEALYRRTITSLVFWKNATKOSTE-EGGQFVTLSPPCP 306
Db 305 LFPYKINY 312
QY 307 EFPYEINY 314

RESULT 4
ID PCT-US99-05485-21 STANDARD; PRT; 312 AA.
XX
AC xxxxxx
DT
XX
DE
XX
Sequence 21, Application PC/TUS9905485
Sequence 21, Application PC/TUS9905485
GENERAL INFORMATION:
```

```
CC APPLICANT: Poet, Steven E.
CC APPLICANT: Ritchie, Branson W.
CC APPLICANT: Niagro, Frank D.
CC APPLICANT: Lukert, Phil D.
CC TITLE OF INVENTION: Vaccines against Circovirus Infections
CC FILE REFERENCE: 21099.0057/P
CC CURRENT APPLICATION NUMBER: PCT/US99/05485
CC CURRENT FILING DATE: 1999-03-12
CC EARLIER APPLICATION NUMBER: 60/077,890
CC EARLIER FILING DATE: 1998-03-13
CC NUMBER OF SEQ ID NOS: 41
CC SOFTWARE: FastSeq for Windows Version 3.0
CC SEQ ID NO 21
CC LENGTH: 312
CC TYPE: PRT
CC ORGANISM: porcine circovirus
CC SEQUENCE 312 AA; 35704 MW; 527440 CN;

Query Match 84.3%; Score 1988; DB 1; Length 312;
Best Local Similarity 84.1%; Pred. No. 1.35e-211;
Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPOPHRWFTLNPNSEKKNIRELPISLFDYFCGEGLEGRPHLQGFANFAK 64
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
8 RSGPOPHRWFTLNPNSEDERKKIRELPISLFDYFVGEEGNEGRPHLQGFANFYK 67
Db 65 QTFNKVWYFGARCHIEKAKGTDOONKEYCSKEGHILIECGAPRNOGRSDLSSTAVSTILL 124
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
68 QTFNKVWYLGARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSOGORSLSSTAVSTILL 127
Db 125 ETGSLVTVAEQFPVTVYRNFRGLAELLKVSCKMOORDWKTAVHVIVGPPGCKSOWARNF 184
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 ESGILTVAEQHPVTFVKNFRGLAELLKVSCKMQRDKNTNVHFIVGPPGCKSKWAANF 187
Db 185 AEPDRTYKPSRNKWDGYHGEVVVLDYFGWLPPWDDLRLCLDRYPLTVETKGTVPFL 244
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 ANPETYKPKPNKWDGYHGEVVVLDYFGWLPPWDDLRLCLDRYPLTVETKGTVPFL 247
Db 245 ARSLITSNAPQEWYSTAPVAEALYRRITTLQFWKTAGEOSTEVEPGEFEAVDPPCA 304
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 ARSLITSNQPLEWYSTAPVAEALYRRITSLVFWKNATKOSTE-BGGQFVTLSPPCP 306
Db 305 LFPYKINY 312
QY :|||:|
307 EFPYEINY 314

RESULT 5
ID PCT-US99-05485-8 STANDARD: PRT; 220 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 8, Application PC/TUS9905485
Sequence 8, Application PC/TUS9905485
GENERAL INFORMATION:
APPLICANT: Poet, Steven E.
APPLICANT: Ritchie, Branson W.
APPLICANT: Niagro, Frank D.
APPLICANT: Lukert, Phil D.
TITLE OF INVENTION: Vaccines against Circovirus Infections
FILE REFERENCE: 21099.0057/P
CURRENT APPLICATION NUMBER: PCT/US99/05485
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: 60/077,890
EARLIER FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 220
TYPE: PRT
CC ORGANISM: beak and feather disease virus
CC SEQUENCE 220 AA; 25485 MW; 260536 CN;

Query Match 27.5%; Score 648; DB 17; Length 220;
Best Local Similarity 45.5%; Pred. No. 4.19e-57;
Matches 96; Conservative 39; Mismatches 68; Indels 8; Gaps 7;

Db 3 PRGHFERAKGSDADNEKYCSKESGDVILITGIVARDGHRH-FDGAVAAVMSGPKMKKEVARE 61
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSOGORSDLSTAVSTLLESGLITVAEQ 138
Db 62 FPDIVVRHGRGLHSLVLSLRP-RDFKTEVDVYIYPPGCGKSRWA-N-EQPTKYYKM- 117
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 HPVTFVKNFRGLAELLKVSCKMQRDKNTNVHFIVGPPGCKSKWAANFANPETTYWKPP 198
Db 118 RGEWMDGYDGEDVWLDYFGWLPCYCEMLRCLDRYPHKVPKGAFAVEFTSKRIIITSNKA 177
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 KKNWMDGYHGEKVVVLDYFGWLPPWDDLRLCLDRYPLTVTKGTVPFLARSILITSNOT 258
Db 178 PETWYKEDCDP--KPLFRFRTR-VWYINIDK 205
QY :|||:|
259 PLEWYSTAPVAEALYRRITSLVFWKNATK 289

Sequence 8, Application US/09267177
Sequence 8, Application US/09267177
GENERAL INFORMATION:
APPLICANT: Poet, Steven E.
APPLICANT: Ritchie, Branson W.
APPLICANT: Niagro, Frank D.
APPLICANT: Lukert, Phil D.
TITLE OF INVENTION: Vaccines against Circovirus Infections
FILE REFERENCE: 21099.0057
CURRENT APPLICATION NUMBER: US/09/267,177
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: 60/077,890
EARLIER FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 220
TYPE: PRT
CC ORGANISM: beak and feather disease virus
CC SEQUENCE 220 AA; 25485 MW; 260536 CN;

Query Match 27.5%; Score 648; DB 17; Length 220;
Best Local Similarity 45.5%; Pred. No. 4.19e-57;
Matches 96; Conservative 39; Mismatches 68; Indels 8; Gaps 7;

Db 3 PRGHFERAKGSDADNEKYCSKESGDVILITGIVARDGHRH-FDGAVAAVMSGPKMKKEVARE 61
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSOGORSDLSTAVSTLLESGLITVAEQ 138
Db 62 FPDIVVRHGRGLHSLVLSLRP-RDFKTEVDVYIYPPGCGKSRWA-N-EQPTKYYKM- 117
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 HPVTFVKNFRGLAELLKVSCKMQRDKNTNVHFIVGPPGCKSKWAANFANPETTYWKPP 198
Db 118 RGEWMDGYDGEDVWLDYFGWLPCYCEMLRCLDRYPHKVPKGAFAVEFTSKRIIITSNKA 177
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 KKNWMDGYHGEKVVVLDYFGWLPPWDDLRLCLDRYPLTVTKGTVPFLARSILITSNOT 258
Db 178 PETWYKEDCDP--KPLFRFRTR-VWYINIDK 205
QY :|||:|
259 PLEWYSTAPVAEALYRRITSLVFWKNATK 289
```

```
CC ORGANISM: beak and feather disease virus
CC SEQUENCE 220 AA; 25485 MW; 260536 CN;

Query Match 27.5%; Score 648; DB 1; Length 220;
Best Local Similarity 45.5%; Pred. No. 4.19e-57;
Matches 96; Conservative 39; Mismatches 68; Indels 8; Gaps 7;

Db 3 PRGHFERAKGSDADNEKYCSKESGDVILITGIVARDGHRH-FDGAVAAVMSGPKMKKEVARE 61
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSOGORSDLSTAVSTLLESGLITVAEQ 138
Db 62 FPDIVVRHGRGLHSLVLSLRP-RDFKTEVDVYIYPPGCGKSRWA-N-EQPTKYYKM- 117
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 HPVTFVKNFRGLAELLKVSCKMQRDKNTNVHFIVGPPGCKSKWAANFANPETTYWKPP 198
Db 118 RGEWMDGYDGEDVWLDYFGWLPCYCEMLRCLDRYPHKVPKGAFAVEFTSKRIIITSNKA 177
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 KKNWMDGYHGEKVVVLDYFGWLPPWDDLRLCLDRYPLTVTKGTVPFLARSILITSNOT 258
Db 178 PETWYKEDCDP--KPLFRFRTR-VWYINIDK 205
QY :|||:|
259 PLEWYSTAPVAEALYRRITSLVFWKNATK 289

RESULT 6
ID US-09-267-177-8 STANDARD: PRT; 220 AA.
XX
AC xxxxxx
XX
DT
XX
DE
XX
Sequence 8, Application US/09267177
Sequence 8, Application US/09267177
GENERAL INFORMATION:
APPLICANT: Poet, Steven E.
APPLICANT: Ritchie, Branson W.
APPLICANT: Niagro, Frank D.
APPLICANT: Lukert, Phil D.
TITLE OF INVENTION: Vaccines against Circovirus Infections
FILE REFERENCE: 21099.0057
CURRENT APPLICATION NUMBER: US/09/267,177
CURRENT FILING DATE: 1999-03-12
EARLIER APPLICATION NUMBER: 60/077,890
EARLIER FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 220
TYPE: PRT
CC ORGANISM: beak and feather disease virus
CC SEQUENCE 220 AA; 25485 MW; 260536 CN;

Query Match 27.5%; Score 648; DB 17; Length 220;
Best Local Similarity 45.5%; Pred. No. 4.19e-57;
Matches 96; Conservative 39; Mismatches 68; Indels 8; Gaps 7;

Db 3 PRGHFERAKGSDADNEKYCSKESGDVILITGIVARDGHRH-FDGAVAAVMSGPKMKKEVARE 61
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
79 ARCHIEKAKGTDOONKEYCSKEGNLLIECGAPRSOGORSDLSTAVSTLLESGLITVAEQ 138
Db 62 FPDIVVRHGRGLHSLVLSLRP-RDFKTEVDVYIYPPGCGKSRWA-N-EQPTKYYKM- 117
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
139 HPVTFVKNFRGLAELLKVSCKMQRDKNTNVHFIVGPPGCKSKWAANFANPETTYWKPP 198
Db 118 RGEWMDGYDGEDVWLDYFGWLPCYCEMLRCLDRYPHKVPKGAFAVEFTSKRIIITSNKA 177
QY :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 KKNWMDGYHGEKVVVLDYFGWLPPWDDLRLCLDRYPLTVTKGTVPFLARSILITSNOT 258
Db 178 PETWYKEDCDP--KPLFRFRTR-VWYINIDK 205
QY :|||:|
259 PLEWYSTAPVAEALYRRITSLVFWKNATK 289
```

RESULT 7
ID US-08-484-735-16 STANDARD; PRT; 286 AA.

XX
AC
XX
DT
XX

Sequence 16, Application US/08484735

Sequence 16, Application US/08484735

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru
APPLICANT: Soong, Tai-Seng
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTION
TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,735
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/418,071
FILING DATE: 06-APR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1 CONT
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 286 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc="DNA I-V1"

SEQUENCE 286 AA; 33108 MW; 434824 CN;

Query Match 8.4%; Score 197; DB 9; Length 286;
Best Local Similarity 36.5%; Pred. No. 3.05e-08;
Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

Db 4 PSLKWCFTLNYSSAAERENLSLLKEEDVHVAVGVDEVPATGOKHLOGYLSLKKRIRLG 63

QY 13 PHKRWFTLNPSEDERKKIRELPISL-FDYFVIGEENEGRTPHLOGFANFYKQTFN 71

Db 64 GLKKKY-GSRAHWEIARGTDEENSKYCS-GTLLLELGFPPVWVNGSNRKKISEMVA 116

QY 72 KVKW-YLGARCHIEKAGTQDNKEYCSKEGNLLIEGAPRSQG-QRSDLSTAYS 124

RESULT 8

ID US-08-484-735-17 STANDARD; PRT; 285 AA.

XX
AC
XX
DT

Sequence 17, Application US/08484735

Sequence 17, Application US/08484735

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru
APPLICANT: Soong, Tai-Seng
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTION
TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,735
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/418,071
FILING DATE: 06-APR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1 CONT
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc="DNA II V2 (D2)"

SEQUENCE 285 AA; 32701 MW; 420559 CN;

Query Match 8.2%; Score 193; DB 9; Length 285;
Best Local Similarity 37.9%; Pred. No. 7.57e-08;
Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 5 KWCFTLNYSSAAEREDFLALLKEEDVHVAVGVDEVPATGOKHLOGYLSLKKRIRLG 64

QY 16 RWVTLNPSEDERKKIRELPISL-FDYFVIGEENEGRTPHLOGFANFYKQTFN 74

Db 65 KKY-GSRAHWEIARGTDEENSKYCSKE-TLVLELGTPTVPVPGSK 105

QY 75 W-YLGARCHIEKAGTQDNKEYCSKEGNLLIEGAPRSQOR 116

RESULT 9

ID US-08-484-735-18 STANDARD; PRT; 286 AA.

XX
AC
XX
DT

Sequence 17, Application US/08484735

Sequence 17, Application US/08484735

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru
APPLICANT: Soong, Tai-Seng
TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETECTION
TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Neave
STREET: 1251 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10020

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 5.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,735
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/418,071
FILING DATE: 06-APR-1995

ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr., James F.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: DCB-1 CONT
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 285 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: /desc="DNA II V2 (D2)"

SEQUENCE 285 AA; 32701 MW; 420559 CN;

Query Match 8.2%; Score 193; DB 9; Length 285;
Best Local Similarity 37.9%; Pred. No. 7.57e-08;
Matches 39; Conservative 18; Mismatches 42; Indels 4; Gaps 4;

Db 5 KWCFTLNYSSAAEREDFLALLKEEDVHVAVGVDEVPATGOKHLOGYLSLKKRIRLG 64

QY 16 RWVTLNPSEDERKKIRELPISL-FDYFVIGEENEGRTPHLOGFANFYKQTFN 74

Db 65 KKY-GSRAHWEIARGTDEENSKYCSKE-TLVLELGTPTVPVPGSK 105

QY 75 W-YLGARCHIEKAGTQDNKEYCSKEGNLLIEGAPRSQOR 116

RESULT 9

ID US-08-484-735-18 STANDARD; PRT; 286 AA.

XX
AC
XX
DT

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

GENERAL INFORMATION:
APPLICANT: Wu, Rey-Yuh
APPLICANT: You, Li-Ru

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

Sequence 18, Application US/08484735

RESULT 14

ID US-08-600-452A-4 STANDARD; PRT; 611 AA.

XX AC xxxxxx

Sequence 4, Application US/08600452A

Sequence 4, Application US/08600452A

GENERAL INFORMATION:

APPLICANT: ROSEMAN, SAUL

APPLICANT: BASSLER, BONNIE

APPLICANT: KEYHANI, NEMAT O.

APPLICANT: CHITLARK, EDITH

APPLICANT: ROWE, CHRIS

APPLICANT: YU, CHARLES

TITLE OF INVENTION: BACTERIAL CATABOLISM OF CHITIN

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSEE: FISH & RICHARDSON P.C.

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/600.452A

FILING DATE: 13-FEB-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Lisa A.

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: 07662/005001

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 611 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 611 AA; 69501 MW; 1856094 CN;

Query Match 4.1%; Score 97; DB 11; Length 611;

Best Local Similarity 25.3%; Pred. No. 4.54e+01;

Matches 19; Conservative 20; Mismatches 31; Indels 5; Gaps 5;

Db 484 WEEAOGDKVSKDTVIYSLWSEQALN-CAROGFDVILQPGQFTYLDIAQDYAPEPGVD 542

Qy 203 WD-GYHGEKVVIDDFYGLWLPWDDLRLCGRYPLTVTKGTGTPVFLARSILITSNQTPL 261

Db 543 W-AGVT-P-LERAYR 554

Qy 262 WYSSTAVPAVEALYR 276

RESULT 15

ID US-60-162-866-574 STANDARD; PRT; 120 AA.

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

XX AC xxxxxx

DE: Sequence 574, Application US/60162866

AX Sequence 574, Application US/60162866

CC GENERAL INFORMATION:

CC APPLICANT: Strabala, Timothy

CC APPLICANT: Nieuwenhuizen, Nicolaas

CC TITLE OF INVENTION: Compositions Isolated from Plant Cells

CC TITLE OF INVENTION: and Their Use in the Modification of Plant Cell Signaling

CC FILE REFERENCE: 11000/1026

CC CURRENT APPLICATION NUMBER: US/60/162,866

CC CURRENT FILING DATE: 1959-11-01

CC NUMBER OF SEQ ID NOS: 1275

CC SOFTWARE: FastSeq for Windows version 3.0

CC SEQ ID NO 574

CC LENGTH: 120

CC TYPE: PRT

CC ORGANISM: Eucalyptus grandis

CC SEQUENCE 120 AA; 13706 MW; 76944 CN;

Query Match 4.0%; Score 94; DB 20; Length 120;

Best Local Similarity 33.8%; Pred. No. 7.86e+01;

Matches 23; Conservative 17; Mismatches 22; Indels 6; Gaps 6;

Db 1 LHRGIEGADSQROSRLYWFAS-AIGAVEVLIFFGWILFKHEV-OSTLEDG-YRATTS 60

Qy 247 LARSIL-ITSN-OTPLEWYSSTAVPAVEAL-YRRITSLVFWKNATKOSTEGGFVTLSP 303

Db 61 QFRFSYE 68

Qy 304 PCPEFPYE 311

Search completed: Wed Dec 22 10:15:36 1999

Time: 41 secs.

WATERMAN

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:13:15 1999; MasPar time 24.23 Seconds
707.243 Million cell updates/sec
Tabular output not generated.

Title: >US-09-209-961-3
Description: (1-314) from US09209961.pep
Perfect Score: 2358
Sequence: 1 MPKNGRSGPQPKRWVFT.....GGQFVTLSPPCPEFFPYEINY 314

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: spstreml9

1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phage 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 46.363; Variance 75.952; scale 0.610

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	2351	99.7	314	14	P35.8.	0.00e+00
2	2296	97.4	314	14	092285	0.00e+00
3	2293	97.2	314	14	093192	0.00e+00
4	1988	84.3	312	14	056123	0.00e+00
5	1988	84.3	312	14	P89032	0.00e+00
6	820	34.8	299	14	090238	0.00e+00
7	242	10.3	290	14	090235	1.32e-167
8	239	10.1	280	14	065005	8.77e-29
9	224	9.5	282	14	087009	3.91e-28
10	221	9.4	284	14	089685	6.40e-25
11	203	8.6	285	14	083030	2.77e-24
12	196	8.3	286	14	083026	1.65e-20
13	136	5.8	286	14	085567	4.58e-19
14	136	5.8	286	14	Q65378	1.96e-07
15	130	5.5	3085	14	070710	1.96e-07
16	120	5.1	144	14	040986	2.28e-06
17	119	5.0	148	14	032663	1.20e-04
18	119	5.0	278	14	068862	1.76e-04
19	111	4.7	102	14	Q65391	3.63e-03
20	112	4.7	135	14	056312	2.50e-03

21 107 4.5 153 14 057304 17.6 KD PROTEIN. 1.57e-02
22 107 4.5 153 14 083478 17.6 KD PROTEIN. 1.57e-02
23 107 4.5 153 14 073572 17.6 KD REPB PROTEIN. 1.57e-02
24 106 4.5 153 14 073477 17.6 KD REPB PROTEIN. 2.25e-02
25 102 4.3 147 14 086645 17.2 KDA ORF. 9.28e-02
26 102 4.3 148 14 067567 5' END OF THE ORF 4 CO 9.28e-02
27 102 4.3 281 14 091252 COMPONENT 9. 9.28e-02
28 102 4.3 290 10 023376 ABC TRANSPORTER HOMOLO 9.28e-02
29 99 4.2 286 14 039828 PUTATIVE REPLICASE-ASS 2.62e-01
30 97 4.1 244 11 070278 MULTIPLE ENDOCRINE NEO 5.17e-01
31 96 4.1 273 10 043381 PSI TYPE III CHLOROPHY 7.23e-01
32 97 4.1 275 8 032304 LIGHT HARVESTING PROTE 5.17e-01
33 97 4.1 611 2 096155 BETA-N-ACETYLHEXOSAMIN 5.17e-01
34 97 4.1 741 2 060081 BETA-N-HEXOSAMINIDASE 5.17e-01
35 96 4.1 1777 14 089278 NONSTRUCTURAL PROTEIN 7.23e-01
36 95 4.0 94 2 096226 HYPOTHETICAL 9.5 KD PR 1.01e+00
37 95 4.0 286 10 020771 CHLOROPHYLL A-B BINDIN 1.01e+00
38 94 4.0 298 10 040839 ETHYLENE-FORMING ENZYM 1.40e+00
39 95 4.0 415 2 006721 RUBULOSE BISPHOSPHATE 1.01e+00
40 94 4.0 825 5 021430 K1E4.4 PROTEIN. 1.40e+00
41 94 4.0 946 2 070888 (BCTA). 1.40e+00
42 92 3.9 183 14 057160 SIMILAR TO AAV2 REP 68 2.69e+00
43 92 3.9 490 5 026855 ALPHA-AMYLASE II. 2.69e+00
44 92 3.9 746 14 067813 VIRAL POLYPROTEIN (FRA 2.69e+00
45 92 3.9 2218 14 067817 COMPLETE GENOME. 2.69e+00

ALIGNMENTS

RESULT 1
ID O92285 PRELIMINARY; PRT; 314 AA.
AC O92285;
DI 01-NOV-1998 (TREMBLREL. 08, CREATED)
FT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE P35.8.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
EA [1]
KP SEQUENCE FROM N.A.
PC STRAIN:412;
RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;
RT "Emergence of a new porcine circovirus."
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF085695; G3668363;
SQ SEQUENCE 314 AA; 35792 MW; F8319091 CRC32;
Query Match 99.7%; Score 2351; DB 14; Length 314;
Pct Idcal Similarity 99.7%; Pred. No. 0.00e+00;
Mismatches 0; Indels 0; Gaps 0;
Cons 313; Conservative 0;

6 1 MPKNGRSGPQPKRWVFTLNNPSEDERKKIRLPISLFDYFIVGEEGEGTPTHLQ 60
7 1 MPKNGRSGPQPKRWVFTLNNPSEDERKKIRLPISLFDYFIVGEEGEGTPTHLQ 60
8 61 FANFVKQTFNKVYLGARCHIEKAGTDQONKEYCSKEGNLLIEGAPRSQGRSDLS 120
9 61 FANFVKQTFNKVYLGARCHIEKAGTDQONKEYCSKEGNLLIEGAPRSQGRSDLS 120
10 121 TAVSTLLESSILTVAKQHPVTFVKNFRLGLAELLKVSQKMRDWTNVHFIVGPGCGK 180
11 121 TAVSTLLESSILTVAKQHPVTFVKNFRLGLAELLKVSQKMRDWTNVHFIVGPGCGK 180
12 181 SKWAANFANPETTYWPKPKWWDGTHGEKVVYIDDFYGLPWDDLRLCDRPLTVTKT 240
13 181 SKWAANFANPETTYWPKPKWWDGTHGEKVVYIDDFYGLPWDDLRLCDRPLTVTKT 240
14 241 GGVVPLARSILITTSNOTPLEWSSYTAAPAVEALYRRTISLVFWKNATKOSTEGGOFVT 300
15 241 GGVVPLARSILITTSNOTPLEWSSYTAAPAVEALYRRTISLVFWKNATKOSTEGGOFVT 300
16 301 LSPPCPEFFPYEINY 314
17 |||||||||||

QY 301 LSPPCPEPPEINY 314

RESULT 2

ID O93192 PRELIMINARY; PRT; 314 AA.
 AC O93192;
 DT 01-NOV-1998 (TREMBREL. 08, CREATED)
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
 DE PUTATIVE REP PROTEIN (ORF1).
 OS PORCINE CIRCOVIRUS.
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RX MEDLINE; 98418498.
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
 RT "Characterisation of novel circovirus DNAs associated with wasting
 RT syndromes in pigs.";
 RL J. GEN. VIROL. 79:2171-2199(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RA MEEHAN B.M.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF055394; G3598832;
 DR EMBL; AF055393; G3598821;
 SQ SEQUENCE 314 AA; 35782 MW; 1C88736B CRC32;

Query Match 97.4%; Score 2296; DB 14; Length 314;

Best Local Similarity 96.8%; Pred. No. 0.00e+00;
Matches 304; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

Db 1 MPKKNRSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOG 60
 QY 1 MPKKNRSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOG 60
 Db 61 FANFVKQTFNKVKYLGARCYIEKAGTDQONKEYCSKEGNLLIECGAPRSQGRSDLS 120
 QY 61 FANFVKQTFNKVKYLGARCYIEKAGTDQONKEYCSKEGNLLIECGAPRSQGRSDLS 120
 Db 121 TAVSTLLESGSLVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 QY 121 TAVSTLLESGSLVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 Db 121 TAVSTLLESGILVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 QY 121 TAVSTLLESGILVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 Db 181 SKWAANFADPETTYKPPKRWKWDGYHGEVVDYFGLPWDDLLRLCDRYPLTVTK 240
 QY 181 SKWAANFADPETTYKPPKRWKWDGYHGEVVDYFGLPWDDLLRLCDRYPLTVTK 240
 Db 241 GGTVPFLARSILITSNPTLEWYSSAVPAVEALYRRTISLVFWKNATEQSTEEGGQFVT 300
 QY 241 GGTVPFLARSILITSNPTLEWYSSAVPAVEALYRRTISLVFWKNATEQSTEEGGQFVT 300
 Db 301 LSPPCPEPPEINY 314
 QY 301 LSPPCPEPPEINY 314

RESULT 3

ID O56123 PRELIMINARY; PRT; 314 AA.
 AC O56123;
 DT 01-JUN-1998 (TREMBREL. 06, CREATED)
 DT 01-JUN-1998 (TREMBREL. 06, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
 DE PUTATIVE REP PROTEIN.
 OS PORCINE CIRCOVIRUS.
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PWVS PCV;
 RX MEDLINE; 98241772.
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;

RT "Nucleotide sequence of porcine circovirus associated with
 RT postweaning multisystemic wasting syndrome in pigs.";
 RL J. VIROL. 72:5262-5267(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PWVS PCV;
 RA HAMEL A.L., LIN L.L., NAYAR G.P.S.;
 RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF027217; G2689646;
 SQ SEQUENCE 314 AA; 35804 MW; 228DA1CE CRC32;

Query Match 97.2%; Score 2293; DB 14; Length 314;

Best Local Similarity 97.1%; Pred. No. 0.00e+00;

Matches 305; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Db 1 MPKKNRSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOG 60
 QY 1 MPKKNRSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOG 60
 Db 61 FANFVKQTFNKVKYLGARCYIEKAGTDQONKEYCSKEGNLLIECGAPRSQGRSDLS 120
 QY 61 FANFVKQTFNKVKYLGARCYIEKAGTDQONKEYCSKEGNLLIECGAPRSQGRSDLS 120
 Db 121 TAVSTLLESGSLVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 QY 121 TAVSTLLESGILVTVAEQHPVTVFNRFRGLAELLKVSQKMKRDKTNNVHVIVGPPCGK 180
 Db 181 SKWAANFADPETTYKPPKRWKWDGYHGEVVDYFGLPWDDLLRLCDRYPLTVTK 240
 QY 181 SKWAANFADPETTYKPPKRWKWDGYHGEVVDYFGLPWDDLLRLCDRYPLTVTK 240
 Db 241 GGTVPFLARSILITSNPTLEWYSSAVPAVEALYRRTISLVFWKNATEQSTEEGGQFVT 300
 QY 241 GGTVPFLARSILITSNPTLEWYSSAVPAVEALYRRTISLVFWKNATEQSTEEGGQFVT 300
 Db 301 LSPPCPEPPEINY 314
 QY 301 LSPPCPEPPEINY 314

RESULT 4

ID P89032 PRELIMINARY; PRT; 312 AA.
 AC P89032;
 DT 01-MAY-1997 (TREMBREL. 03, CREATED)
 DT 01-MAY-1997 (TREMBREL. 03, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBREL. 08, LAST SEQUENCE UPDATE)
 DE PUTATIVE PCV REPLICATION-ASSOCIATED PROTEIN (REP).
 OS PORCINE CIRCOVIRUS.
 OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE; 97163487.
 RA MEEHAN B.M., CREELAN J.L., MCNUITY S.M., TODD D.;
 RT "Sequence of porcine circovirus DNA: affinities with plant
 RT circoviruses.";
 RL J. GEN. VIROL. 78:221-227(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RA MEEHAN B.M.;
 RL SUBMITTED (FEB-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; U49186; G1841515;
 DR PFAM; PF00799; Geminivirus_all; 1.
 SQ SEQUENCE 312 AA; 35704 MW; 7434D27B CRC32;

Query Match 84.3%; Score 1988; DB 14; Length 312;

Best Local Similarity 84.1%; Pred. No. 0.00e+00;

Matches 259; Conservative 22; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOGFANFAK 64
 QY 8 RSGPQPHKRWVFTLNPNSEDERKKIRDLPTSLFDYFVIGEGNEGRTPHLOGFANFAK 67

```

Db 65 QTFNKVYFGARCHEIAKAGTDOONKEYCSKEGHTLIECGAPRNOGKRSDDLSTAVSTILL 124
Qy 68 QTFNKVYGLGARCHIEKAGTDOONKEYCSKEGNLLIECGAPRNOGKRSDDLSTAVSTILL 127
Db 125 ETGSLVTVAEQPFVTVNRFRGLAELLKVSQKQORDWKTAVHVIIVGPGCGKSQWARNF 184
Qy 128 ESGILVTVAEQPFVTVNRFRGLAELLKVSQKQORDWKTAVHVIIVGPGCGKSQWARNF 187
Db 185 APRDITWYKPSRNKMDGVBHGEVVDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 244
Qy 188 ANPETTYKPPKKNWMDGVBHGEVVDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 247
Db 245 ARSILITSNQAPQEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRREAVDPPCA 304
Qy 248 ARSILITSNQPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRREAVDPPCA 306
Db 305 LPPYKINY 312
Qy 307 EPPYEINY 314

RESULT 5
ID O90238 PRELIMINARY; PRT; 312 AA.
AC O90238;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE REPLICATION-ASSOCIATED PROTEIN.
GN V1.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99003925.
RA NIAGRO F.D., FORTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
RT intermediates between the geminiviruses and plant circoviruses."
RT ARCH. VIROL. 143:1723-1744 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF071879; G3551844;
FT NON-TER 1
FT SEQUENCE 312 AA; 35665 MW; 8138CD07 CRC32;
SQ

Query Match 84.3%; Score 1988; DB 14; Length 312;
Best Local Similarity 83.8%; Pred. No. 0.00e+00;
Matches 258; Conservative 23; Mismatches 26; Indels 1; Gaps 1;

Db 5 KSGPQPHKRWVTLNPPSEENKRELPLISLFDYFVCGEBLEGRTHLQGFANFAK 64
Qy 8 RSGPQPHKRWVTLNPPSEENKRELPLISLFDYFVCGEBLEGRTHLQGFANFAK 67
Db 65 QTFNKVYFGARCHEIAKAGTDOONKEYCSKEGHTLIECGAPRNOGKRSDDLSTAVSTILL 124
Qy 68 QTFNKVYGLGARCHIEKAGTDOONKEYCSKEGNLLIECGAPRNOGKRSDDLSTAVSTILL 127
Db 125 ETGSLVTVAEQPFVTVNRFRGLAELLKVSQKQORDWKTAVHVIIVGPGCGKSQWARNF 184
Qy 128 ESGILVTVAEQPFVTVNRFRGLAELLKVSQKQORDWKTAVHVIIVGPGCGKSQWARNF 187
Db 185 TEPDITWYKPSRNKMDGVBHGEVVDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 244
Qy 188 ANPETTYKPPKKNWMDGVBHGEVVDYFGWLPWDDLRLCDRLPLVETKGGTVPEL 247
Db 245 ARSILITSNQAPQEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRREAVDPPCA 304
Qy 248 ARSILITSNQPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRREAVDPPCA 306
Db 305 LPPYKINY 312
Qy 307 EPPYEINY 314

```

```

Qy 307 EPPYEINY 314

RESULT 6
ID O90235 PRELIMINARY; PRT; 299 AA.
AC O90235;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE REPLICATION-ASSOCIATED PROTEIN (FRAGMENT).
GN V1.
OS BEAK AND FEATHER DISEASE VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99003925.
RA NIAGRO F.D., FORTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
RT intermediates between the geminiviruses and plant circoviruses."
RT ARCH. VIROL. 143:1723-1744 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORTHOEFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL: AF071878; G3551840;
FT NON-TER 1
FT SEQUENCE 299 AA; 34479 MW; 248AAC02 CRC32;
SQ

Query Match 34.8%; Score 820; DB 14; Length 299;
Best Local Similarity 45.5%; Pred. No. 1.32e-167;
Matches 125; Conservative 50; Mismatches 90; Indels 10; Gaps 9;

Db 20 RWCFTLNPTDGEIEFVRSIGDPDEFYAIYVGRKBDQ-TPHLOGFHFHKNKRLSALK 78
Qy 15 KRWFVTLNPPSEDERKKRELPLISLFDYFVCGEBLEGRTHLQGFANFAKQTFNKV 74
Db 79 KML-PRGHEPERAKGSDADNEKCYSGEDVILTLGIVARDGHR-FCGAVAAVMSPKME 136
Qy 75 WYLGARCHEIAKAGTDOONKEYCSKEGNLLIECGAPRNOGKRSDDLSTAVSTILL 134
Db 137 VAREFPDIYVRHGRGLSLLSLVGSRP-RDFKTEVDYIYVPGCGKSRWA-N-QPQGTGY 193
Qy 135 VAOQHPVTFVKNFRGLAELLKVSQKQORDWKTAVHVIIVGPGCGKSQWARNF 194
Db 194 YKM-RGEWMDGVDGDDVILDDYFGLWLPYCEMLRLCDRLPLVETKGGTVPEL 252
Qy 195 WKPPKKNWMDGVBHGEVVDYFGWLPWDDLRLCDRLPLVETKGGTVPELARSILIT 254
Db 253 SNKAPETWTKEDCDP-KPLFRFRTR-VWYNYNDK 284
Qy 255 SNOTPLEWYSSSTAVPAVEALYRITITLQFWKTAGSQSTVEPGRREAVDPPCA 289

RESULT 7
ID Q66005 PRELIMINARY; PRT; 290 AA.
AC Q66005;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ORF 1.
OS COCONUT FOLIAR DECAY VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90266484.
RA KOHDE W., RANDES J.W., LANGRIDGE P., HANOLD D.;
RT "Nucleotide sequence of a circular single-stranded DNA associated
RT with coconut foliar decay virus."
RT VIROLOGY 176:648-651 (1990).
DR EMBL: M29963; G323307;
SQ

```

```

SQ SEQUENCE 290 AA; 33497 MW; 87A78804_CRC32;

Query Match 10.3%; Score 242; DB 14; Length 290;
Best Local Similarity 38.2%; Pred. No. 8.77e-29;
Matches 42; Conservative 26; Mismatches 36; Indels 6; Gaps 6;

Db 6 RNCFTINTETEEAANVVRRIEISLNV-YAIVGDEVAPSTGQRHLQGFPHLKTGRRLOG 64
QY 15 KRWFFTLNPPSEDERKKI-REL-PISLFDFYFVIGEEGNEGRTPHLQGFANFVKQTFFNK 72
Db 65 LKTVLGNDRHLPTRGSDQONRDYSKE-RVLLHGVTPRCVKPRRLA 113
QY 73 KRWYLG-A-RCHIEKAKGTDOONKEYSKEGNLLIEGAPRSOG-QRSDLS 120

RESULT 8
ID O87009; PRELIMINARY; PRT; 280 AA.
AC O87009;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POSSIBLE REPLICATION ASSOCIATED PROTEIN (SCSV2).
DE SCSV2.
OS SUBTERRANEAN CLOVER STUNT VIRUS.
OC VIRUSES; SS DNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 95193233.
RA BOEVINK P., CHU P.W., KEESE P.;
RT "Sequence of subterranean clover stunt virus DNA: affinities with the
RT geminiviruses.";
RL Virology 207:354-361(1995).
SQ EMBL; U16731; G571488; -.
SR SEQUENCE 280 AA; 32473 MW; 3A60B6C7_CRC32;

Query Match 10.1%; Score 239; DB 14; Length 280;
Best Local Similarity 41.2%; Pred. No. 3.91e-28;
Matches 40; Conservative 20; Mismatches 34; Indels 3; Gaps 3;

Db 3 RRYCFTLNYATERETFTSLFSQDELINTFVVGDEFTATTGQK-HLQGFYSFKNKIRLGL 61
QY 15 KRWFFTLNPPSEDERKKIREL-PISLFDFYFVIGEEGNEGRTPHLQGFANFVKQTFFNK 73
Db 62 KKAFGNRAHWETARGSDQNRDYCKE-TLISIGIP 97
QY 74 KWLFGARCHIERAKGTDOONKEYSKEGNLLIECGAP 110

RESULT 9
ID O91254 PRELIMINARY; PRT; 282 AA.
AC O91254;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE COMPONENT 1-EG.
DE CI-EG.
OS FABIA BEAN NECROTIC YELLOWS VIRUS.
OC VIRUSES; SS DNA VIRUSES; CIRCOVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-EVI-93;
RA KATUL L., TIMCHENKO T., GRONENBORN B., VETTEN H.J.;
RT "The faba bean necrotic yellows virus genome comprises at least ten
RT circular ssDNA components four of which encode putative Rep
RT proteins."
RL SUBMITTED (MAY-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
SQ EMBL; AU005968; E1318733; -.
SR SEQUENCE 282 AA; 32676 MW; 9D32A16E_CRC32;

Query Match 9.5%; Score 224; DB 14; Length 282;
Best Local Similarity 39.6%; Pred. No. 6.40e-25;
Matches 40; Conservative 18; Mismatches 41; Indels 2; Gaps 2;

```

DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST ANNOTATION UPDATE)
 DE (BBTV DNA I) VI AND CI-C3 GENES, COMPLETE CDS S.
 GN VI.
 OS BANANA BUNCHY TOP VIRUS.
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA WU R.Y., YOU L.R., SOONG T.S.;
 RL SUBMITTED (AUG-1994) TO EMBL/GENBANK/DDBJ DATA BANKS.
 DR EMBL; L32166; G520787;
 SQ SEQUENCE 286 AA; 33249 MW; 52927C5B CRC32;

Query Match 8.3%; Score 196; DB 14; Length 286;
 Best Local Similarity 36.5%; Pred. No. 4.58e-19;
 Matches 42; Conservative 21; Mismatches 47; Indels 5; Gaps 5;

Db 4 PSLKWCFTLNYSSAARENELSLKEEDVHVAVGVDEVPATGQKHLOGYLSLKKRIRLG 63
 QY 13 PHKRWVFTLNPNSEDERKKIRELPISLFDYFVGEENEGRTPHLOGFANFVKQTFN 71
 Db 64 GLKKKY-GSRAWEIARTDEBNSKYCSKE-TLILELGFPPVVGNSKKRISMWVA 116
 QY 72 KVKW-YLGARCHIEKAKTDOONKEYCSKEGNLLIECCAPRSOG-QRSDLSTAVS 124

RESULT 13
 ID Q86567 PRELIMINARY; PRT; 286 AA.
 AC Q86567;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE ORF V1 [COMPONENT 1].
 OS BANANA BUNCHY TOP VIRUS.
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 93187592.
 RA HARDING R.M., BURNS T.M., HAFNER G., DIETZGEN R.G., DALE J.L.;
 RT "Nucleotide sequence of one component of the banana bunchy top virus
 genome contains a putative replicase gene."
 RL J. GEN. VIROL. 74:323-328(1993).
 DR EMBL; S56276; G298334;
 SQ SEQUENCE 286 AA; 33603 MW; C2FBE5C7 CRC32;

Query Match 5.8%; Score 136; DB 14; Length 286;
 Best Local Similarity 31.0%; Pred. No. 1.96e-07;
 Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMFTINNP-TLPV-MRD-EIK-YMYOV-ERG-QEG-TRHVQGYVEMKRSSLKQMRGF 60
 QY 17 WFTLNPNSEDERKKIRELPISLFDYFVGEENEGRTPHLOGFANFVKQTFN 76
 Db 61 PGA-HLEKRGSOEARSYCKE 83
 QY 77 LGARCHIEKAKTDOONKEYCSKE 100

RESULT 14
 ID Q65378 PRELIMINARY; PRT; 286 AA.
 AC Q65378;
 DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
 DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 33.5 KD PROTEIN.
 OS BANANA BUNCHY TOP VIRUS.
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HAWAIIAN;
 RA XIE W.S., HU J.S.;
 RL PHYTOPATHOLOGY 0:0-0(0).
 DR EMBL; U18077; G603196;

KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 286 AA; 33488 MW; 3C722F3E CRC32;
 Query Match 5.8%; Score 136; DB 14; Length 286;
 Best Local Similarity 31.0%; Pred. No. 1.96e-07;
 Matches 26; Conservative 27; Mismatches 23; Indels 8; Gaps 8;

Db 8 WMFTINNP-TLPV-MRD-EIK-YMYOV-ERG-QEG-TRHVQGYVEMKRSSLKQMRGF 60
 QY 17 WFTLNPNSEDERKKIRELPISLFDYFVGEENEGRTPHLOGFANFVKQTFN 76
 Db 61 PGA-HLEKRGSOEARSYCKE 83
 QY 77 LGARCHIEKAKTDOONKEYCSKE 100

RESULT 15
 ID Q70710 PRELIMINARY; PRT; 3085 AA.
 AC Q70710;
 DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
 DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
 DE POLYPROTEIN.
 OS INFECTIOUS FLACHERIE VIRUS.
 OC VIRUSES; SSRNA POSITIVE-STRAND VIRUSES, NO DNA STAGE; PICORNAVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98166871.
 RA ISAWA H., ASANO S., SAHARA K., IIZUKA T., BANDO H.;
 RT "Analysis of genetic information of an insect picorna-like virus,
 infectious flacherie virus of silkworm: evidence for evolutionary
 RT relationships among insect, mammalian and plant picorna(-like)
 RT virus-s."
 RL ARCH. VIROL. 143:127-143(1998).
 DR EMBL; AB000906; D1026299;
 KW POLYPROTEIN.
 SQ SEQUENCE 3085 AA; 345798 MW; EEB514A1 CRC32;

Query Match 5.5%; Score 130; DB 14; Length 3085;
 Best Local Similarity 35.3%; Pred. No. 2.28e-06;
 Matches 12; Conservative 13; Mismatches 9; Indels 0; Gaps 0;

Db 1411 PGKFEHPNPVYTRNPTSONWNGYNGOPIVLYDDI 1444
 QY 184 AAFANPETTYKPKNKWWDYHGEKVVVIDDF 217

Search completed: Wed Dec 22 10:13:43 1999
 Job time : 28 secs.

 W P E S R L H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:29:55 1999; MasPar time 9.99 Seconds
 Tabular output not generated. 289.470 Million cell updates/sec

Title: >US-09-209-961-21
 Description: (1-53) from US09209961.pep
 Perfect Score: 416
 Sequence: 1 MVFIIHLGFKWGVFKIFSE.....VFTVFERSAEAYVHHISRGL 53

Scoring table: PAM 150
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl9

1:sp.archaea 2:sp.bacteria 3:sp.fungi 4:sp.human
 5:sp.invertebrate 6:sp.mammal 7:sp.mhc 8:sp.organelle
 9:sp.phase 10:sp.plant 11:sp.rodent 12:sp.unclassified
 13:sp.vertebrate 14:sp.virus

Statistics: Mean 37.523; Variance 83.063; scale 0.452

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	416	100.0	53	14	P6.2.	4.88e-54
2	412	99.0	53	14	ORF5.	2.55e-53
3	406	97.6	53	14	STRAIN PMWS PCV, COMPL	3.04e-52
4	95	22.8	124	2	HYPOHETICAL 14.5 KD P	3.88e-01
5	92	22.1	1199	14	DNA-DEPENDENT RNA POLY	9.18e-01
6	91	21.9	292	2	ABC TRANSPORTER.	1.22e+00
7	90	21.6	236	8	NADH DEHYDROGENASE SUB	1.62e+00
8	90	21.6	236	8	NADH DEHYDROGENASE SUB	1.62e+00
9	89	21.4	302	2	PROBABLE TRANSMEMBRANE	2.14e+00
10	87	20.9	276	2	ABC TRANSPORTER.	3.72e+00
11	87	20.9	1268	5	FGP-3 PROTEIN.	3.72e+00
12	86	20.7	337	5	T23D5.6 PROTEIN.	4.90e+00
13	86	20.7	434	10	F1104.6 PROTEIN.	4.90e+00
14	85	20.4	315	1	HYPOHETICAL 36.7 KD P	6.43e+00
15	85	20.4	377	5	R11D1.5 PROTEIN.	6.43e+00
16	84	20.2	152	8	NADH DEHYDROGENASE SUB	8.43e+00
17	84	20.2	325	8	NADH DEHYDROGENASE SUB	8.43e+00
18	84	20.2	422	2	FEMA.	8.43e+00
19	84	20.2	438	10	ALPHA-AMYLASE 2.	8.43e+00
20	84	20.2	477	3	CHROMOSOME XII COSMID	8.43e+00

21 84 20.2 924 2 006842 B14, PUTATIVE POLYGALA 8.43e+00
 22 83 20.0 106 1 006101 HYPOHETICAL 12.8 KD P 1.10e+01
 23 83 20.0 120 2 068285 LARGE CONDUCTANCE MECH 1.10e+01
 24 83 20.0 236 8 063674 NADH DEHYDROGENASE SUB 1.10e+01
 25 83 20.0 236 8 063671 NADH DEHYDROGENASE SUB 1.10e+01
 26 83 20.0 236 8 063944 NADH DEHYDROGENASE SUB 1.10e+01
 27 83 20.0 236 8 063670 NADH DEHYDROGENASE SUB 1.10e+01
 28 83 20.0 236 8 063666 NADH DEHYDROGENASE SUB 1.10e+01
 29 83 20.0 239 5 044810 F14D2.1 PROTEIN. 1.10e+01
 30 82 19.7 100 3 P87321 ORF YJ025C-A. 1.44e+01
 31 82 19.7 152 8 096091 NADH DEHYDROGENASE SUB 1.44e+01
 32 82 19.7 152 8 095830 NADH DEHYDROGENASE SUB 1.44e+01
 33 82 19.7 209 2 P95773 CADB. 1.44e+01
 34 82 19.7 236 8 063969 NADH DEHYDROGENASE SUB 1.44e+01
 35 82 19.7 236 8 063668 NADH DEHYDROGENASE SUB 1.44e+01
 36 82 19.7 299 5 P91070 SIMILARITY TO C. ELEGA 1.44e+01
 37 82 19.7 383 1 026855 CONSERVED PROTEIN. 1.44e+01
 38 82 19.7 489 5 093704 F42D1.3 PROTEIN. 1.44e+01
 39 82 19.7 528 10 080838 PUTATIVE ATP-DEPENDENT 1.44e+01
 40 82 19.7 840 3 005848 J1575L ORF. 1.44e+01
 41 82 19.7 1435 11 088922 CL2BB. 1.44e+01
 42 82 19.7 1452 11 088918 CL2AA. 1.44e+01
 43 82 19.7 1457 11 088921 CL2BA. 1.44e+01
 44 82 19.7 1478 11 088923 CL2BC. 1.44e+01
 45 82 19.7 2035 4 092508 MYELOBLAST KIAA0233. 1.44e+01

ALIGNMENTS

RESULT 1
 ID O93201 PRELIMINARY; PRT; 53 AA.
 AC O93201;
 AT 01-NOV-1999 (TREMREL. 08, CREATED)
 CT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE P6.2.
 OS PORCINE CIRCOVIRUS.
 OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-412;
 RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;
 RT "Emergence of a new porcine circovirus."
 RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 VN [2]
 XP SEQUENCE FROM N.A.
 FC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RX MEDLINE: 98418498.
 RA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWHRST V.,
 RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
 RT "Characterisation of novel circovirus DNAs associated with wasting
 syndromes in pigs."
 RL J. GEN VIROL. 79:2171-2199(1998).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
 RA MEEHAN B.M.;
 RI SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 BR EMBL: AF085695; G3668366;
 FR EMBL: AF055392; G3598813;
 SQ SEQUENCE 53 AA; 6176 MW; 59859B1D CRC32;
 Query Match 100.08; Score 416; DB 14; Length 53;
 Desc Local Similarity 100.08; Pred. No. 4.88e-54;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 MVFIIHLGFKWGVFKIFSELYIHGTDIVLVVFTVFERSAEAYVHHISRGL 53
 QY 1 MVFIIHLGFKWGVFKIFSELYIHGTDIVLVVFTVFERSAEAYVHHISRGL 53
 RESULT 2
 ID O91861 PRELIMINARY; PRT; 53 AA.

RESULT 10
ID P72880 PRELIMINARY; PRT; 276 AA.
AC P72880;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE ABC TRANSPORTER.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA KANEKO T., SATO S., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA MIYAJIMA N., HIROSAWA M., SUGIURA M., SASAMOTO S., KIMURA T.,
RA HOSOUCHI T., MATSUNO A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YANADA M., YASUDA M.,
RA TABATA S.;
RT Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT genome and assignment of potential protein-coding regions.*;
RL DNA RES. 3:109-136(1996).
DR EMBL; D90901; D1017629.
DR PFAM; PF01061; ABC2_membrane; 1.
SQ SEQUENCE 276 AA; 31542 MW; 4DBF5330 CRC32;

Query Match 20.9%; Score 87; DB 2; Length 276;
Best Local Similarity 28.9%; Pred. No. 3.72e+00;
Matches 11; Conservative 14; Mismatches 10; Indels 3; Gaps 3;

Db 224 MVSVID-GFMAILGSE-STIFLPGFL-LSLLLVIIIF 258
QY 1 MVFIIHLGFKWGVKIKFSELYTHGYTDIVLVVFTVF 38

RESULT 11
ID Q20335 PRELIMINARY; PRT; 1268 AA.
AC Q20335;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE PGP-3 PROTEIN.
GN PGP-3.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA WHITE S.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z82051; E1354574;
SQ SEQUENCE 337 AA; 39448 MW; AD531276 CRC32;

Query Match 20.7%; Score 86; DB 5; Length 337;
Best Local Similarity 25.5%; Pred. No. 4.90e+00;
Matches 13; Conservative 17; Mismatches 18; Indels 3; Gaps 2;

Db 336 VLYAINFGFQWALGLYKLNVEDQYSLYMRQVLVDYDVF-DISQVPCVIHV 185
QY 1 MVFIIHLGFKW-GVFKIKFSELYTHGYTDIVLVVFTVFSAEAYVVIH 49

RESULT 13
ID O82589 PRELIMINARY; PRT; 434 AA.
AC O82589;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F1104:6 PROTEIN.
GN F1104:6.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CROSS).

RN [3] SEQUENCE FROM N.A.
RP LENNARD N.;
RL SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z66567; E1351093;
DR EMBL; Z66562; E1351093; JOINED.
DR EMBL; Z66562; E1346669;
DR EMBL; Z66567; E1346669; JOINED.
DR PROSITE; PS00211; ABC_TRANSPORTER; 2.
KW ATP-BINDING; TRANSPORT.
SQ SEQUENCE 1268 AA; 140290 MW; 48088CF6 CRC32;

Query Match 20.9%; Score 87; DB 5; Length 1268;
Best Local Similarity 36.4%; Pred. No. 3.72e+00;
Matches 8; Conservative 6; Mismatches 8; Indels 0; Gaps 0;

Db 712 LVFTLIRGTWPAFSIVYGOLF 733
QY 1 MVFIIHLGFKWGVKIKFSELY 22

RESULT 12
ID O45805 PRELIMINARY; PRT; 337 AA.
AC O45805;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE T23D5:6 PROTEIN.
GN T23D5:6.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEIA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RA LLOYD C., WILKINSON J.;
RL SUBMITTED (NOV-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., COULSON A.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., FULTON L.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SWALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL NATURE 368:32-38(1994).
DR EMBL; Z82051; E1354574;
SQ SEQUENCE 337 AA; 39448 MW; AD531276 CRC32;

Query Match 20.7%; Score 86; DB 5; Length 337;
Best Local Similarity 25.5%; Pred. No. 4.90e+00;
Matches 13; Conservative 17; Mismatches 18; Indels 3; Gaps 2;

Db 336 VLYAINFGFQWALGLYKLNVEDQYSLYMRQVLVDYDVF-DISQVPCVIHV 185
QY 1 MVFIIHLGFKW-GVFKIKFSELYTHGYTDIVLVVFTVFSAEAYVVIH 49

RESULT 13
ID O82589 PRELIMINARY; PRT; 434 AA.
AC O82589;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE F1104:6 PROTEIN.
GN F1104:6.
OS ARABIDOPSIS THALIANA (MOUSE-EAR CROSS).

 W P S R L H

 (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:17:08 1999; MasPar time 19.80 Seconds
 642.264 Million cell updates/sec
 Tabular output not generated.

Title: >US-09-209-961-5
 Description: (1-233) from US09209961.pep
 Perfect Score: 1805
 Sequence: 1 MTYPRRRRRRRRPRSHLG.....RVMTYVQFREFNLKDPPLP 233

Scoring table: PAM 150
 Gap 11

Searched: 179066 seqs, 54579741 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: sptrembl9

1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
 5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
 9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
 13:sp_vertebrate 14:sp_virus

Statistics: Mean 45.232; Variance 85.435; scale 0.529

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	1795	99.4	233	14	ORF2.	0.00e+00
2	1746	96.7	233	14	ORF2.	0.00e+00
3	1746	96.7	233	14	STRAIN PMWS PCV, COMPL	0.00e+00
4	1729	95.8	233	14	ORF2.	0.00e+00
5	1196	66.3	230	14	CAPSID PROTEIN.	1.65e-228
6	233	12.9	244	14	CAPSID PROTEIN (FRAGME	2.59e-23
7	146	8.1	770	14	LONG ORF.	6.62e-08
8	143	7.9	770	14	LONG ORF.	2.03e-07
9	143	7.9	770	14	ORF1.	2.03e-07
10	140	7.8	768	14	LONG ORF.	6.17e-07
11	139	7.7	770	14	ORF1 PROTEIN.	8.91e-07
12	133	7.4	769	14	LONG ORF.	7.91e-06
13	133	7.4	769	14	LONG ORF.	7.91e-06
14	133	7.4	769	14	LONG ORF.	7.91e-06
15	133	7.4	770	14	LONG ORF.	7.91e-06
16	127	7.0	124	14	LONG ORF. (FRAGMENT).	6.72e-05
17	125	6.9	769	14	LONG ORF.	1.36e-04
18	120	6.6	134	2	HYPOTHETICAL 15.1 KD P	7.66e-04
19	120	6.6	1538	2	PGRS-FAMILY PROTEIN.	7.66e-04
20	115	6.4	823	11	PRP4 PROTEIN KINASE HO	4.16e-03

21 113 6.3 62 6 018749 PROTAMINE P1. 8.11e-03
 22 114 6.3 113 3 008428 ORF FOR053W. 5.82e-03
 23 110 6.1 126 3 012444 ORF FOR309C. 2.17e-02
 24 110 6.1 586 11 P97358 TAFI68. 2.17e-02
 25 109 6.0 225 1 058203 225AA LONG HYPOTHETICA 3.01e-02
 26 106 5.9 171 14 071097 PVII. 7.90e-02
 27 104 5.8 200 14 098691 COAT PROTEIN. 1.49e-01
 28 104 5.8 327 5 002394 DROSOPHILA MELANOGAST 1.49e-01
 29 103 5.7 344 5 016526 T05H4.1 PROTEIN. 2.04e-01
 30 102 5.7 449 14 096673 NUCLEOCAPSID PROTEIN V 2.79e-01
 31 102 5.7 1288 4 015038 KIAA0324 (FRAGMENT). 2.79e-01
 32 102 5.7 1791 4 060382 KIAA0324 (FRAGMENT). 2.79e-01
 33 101 5.6 271 10 023630 TRANSCRIPTION FACTOR. 3.80e-01
 34 101 5.6 710 3 000210 CHROMOSOME SEGREGATION 3.80e-01
 35 101 5.6 968 4 060289 KIAA0542 PROTEIN. 3.80e-01
 36 101 5.6 1473 11 063625 CTD-BINDING SR-LIKE PR 3.80e-01
 37 100 5.5 111 10 039682 GLYCINE-RICH PROTEIN (5.17e-01
 38 99 5.5 153 10 039680 HEAT-SHOCK COGNATE (FR 7.02e-01
 39 100 5.5 163 4 015410 NUCLEIC ACID BINDING P 5.17e-01
 40 95 5.5 302 3 005330 PUTATIVE ORF. 7.02e-01
 41 100 5.5 449 14 096672 COMPLETE GENOME: VP2, 5.17e-01
 42 100 5.5 449 14 041449 52 KDA MAJOR CAPSID PR 5.17e-01
 43 99 5.5 474 2 P75295 P02_ORF474 PROTEIN. 7.02e-01
 44 98 5.4 249 2 083299 UDP-N-ACETYL-D-MANNOSA 9.51e-01
 45 98 5.4 2183 14 086887 POLYPROTEIN. 9.51e-01

ALIGNMENTS

RESULT 1
 ID 091865 PRELIMINARY; PRT; 233 AA.
 AC 091865;
 DT 01-NOV-1998 (TREMREL. 08, CREATED)
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
 DE ORF2.
 OS PORCINE CIRCOPOLVIRUS.
 CC VIRUSES; SSDNA VIRUSES; CIRCOPOLVIRIDAE; CIRCOPOLVIRUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOPOLVIRUS TYPE II;
 RX MEDLINE; 98418498.
 PA MEEHAN B.M., MCNEILLY F.M., TODD D., KENNEDY S., JEWELL V.,
 ELIAS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
 "Characterisation of novel circovirus DNAs associated with wasting
 syndromes in pigs."; J. GEN. VIROL. 79:2171-2159(1998).
 J. GEN. VIROL. 79:2171-2159(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-PORCINE CIRCOPOLVIRUS TYPE II;
 RA MEEHAN B.M.;
 RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
 DR EMBL; AF35392; G3598810;
 EQ SEQUENCE 233 AA; 27840 MW; 721197FF CRC32;
 Query Match 99.4%; Score 1795; DB 14; Length 233;
 Best Local Similarity 99.1%; Pred. No. 0.00e+00;
 Matches 231; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 LQ 1 MTYPRRRRRRRRPRSHLGQLRRPWLPHRPHRYRWRKNGIFNTRLSRTFGYTVKRT 60
 QY 1 MTYPARRRRRRPRSHLGQLLRRLPWLPHRPHRYRWRKNGIFNTRLSRTFGYTVKRT 60
 DF 61 TVTTPSWAVDMRFRKIDDPVPPGGGKNIKIPFEYIRIRKRVKVEFWPCSPITQDGRVGS 120
 QY 61 TVTTPSWAVDMRFRKIDDPVPPGGGKNIKIPFEYIRIRKRVKVEFWPCSPITQDGRVGS 120
 DF 121 TAVILDDNFVTATATYDPYNYSSRRHTIPQPFSSHSRYFTPKPVLDSIDTFQPNKR 180
 QY 121 TAVILDDNFVTATATYDPYNYSSRRHTIPQPFSSHSRYFTPKPVLDSIDTFQPNKR 180
 DF 181 NOLWLRLOTSGNVDHVGCGAAAFENSKYDODYNIIRVTMYVQFREFNLKDPPLP 233
 QY 181 NOLWLRLOTSGNVDHVGCGAAAFENSKYDODYNIIRVTMYVQFREFNLKDPPLP 233

QY 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

RESULT 2

ID O91862 PRELIMINARY; PRT; 233 AA.
AC O91862;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF2.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RX MEDLINE; 98418498.
RA MEEHAN B.M.; MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
RT "Characterisation of novel circovirus DNAs associated with wasting
RT syndromes in pigs.";
RL J. GEN. VIROL. 79:2171-2199(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RA MEEHAN B.M.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF055391; G3598798;
SQ SEQUENCE 233 AA; 27882 MW; 8C4A18BA CRC32;

Query Match 96.7%; Score 1746; DB 14; Length 233;
Best Local Similarity 97.0%; Pred. No. 0.00e+00;
Matches 226; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKAT 60
QY 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKRT 60
Db 61 TVRTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
QY 61 TVTTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
Db 121 TAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
QY 121 TAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
Db 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLK 233
QY 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

RESULT 3

ID O56129 PRELIMINARY; PRT; 233 AA.
AC O56129;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE STRAIN PMWS PCV, COMPLETE GENOME.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PMWS PCV;
RX MEDLINE; 98241772.
RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;
RT "Nucleotide sequence of porcine circovirus associated with
RT postweaning multisystemic wasting syndrome in pigs.";
RL J. VIROL. 72:5262-5267(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-PMWS PCV;
RA HAMEL A.L., LIN L.L., NAVAR G.P.S.;
RL SUBMITTED (SEP-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF027217; G2689647;

SO SEQUENCE 233 AA; 27897 MW; CB1F261D CRC32;

Query Match 96.7%; Score 1746; DB 14; Length 233;
Best Local Similarity 97.0%; Pred. No. 0.00e+00;
Matches 226; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKAT 60
QY 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKRT 60
Db 61 TVRTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
QY 61 TVTTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
Db 121 TAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
QY 121 TAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
Db 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLK 233
QY 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

RESULT 4

ID O91868 PRELIMINARY; PRT; 233 AA.
AC O91868;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE ORF2.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RX MEDLINE; 98418498.
RA MEEHAN B.M.; MCNEILLY F.M., TODD D., KENNEDY S., JEWURST V.,
RA ELLIS J.A., HASSARD L.E., CLARK E.G., HAINES D.M., ALLAN G.M.;
RT "Characterisation of novel circovirus DNAs associated with wasting
RT syndromes in pigs.";
RL J. GEN. VIROL. 79:2171-2199(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-PORCINE CIRCOVIRUS TYPE II;
RA MEEHAN B.M.;
RL SUBMITTED (MAR-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF055393; G3598822;
SQ SEQUENCE 233 AA; 27867 MW; A6EBC19 CRC32;

Query Match 95.8%; Score 1729; DB 14; Length 233;
Best Local Similarity 94.0%; Pred. No. 0.00e+00;
Matches 219; Conservative 9; Mismatches 5; Indels 0; Gaps 0;

Db 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKRT 60
QY 1 MTPRRYRRRRHPRSHLQGLRRPWLHPHRYRWRKNGIFNRLSRTFGYTVKRT 60
Db 61 TVRTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
QY 61 TVTTPSWAVDMRNFIDFPVGGTGNKISIPPEYIRKVKVEFWPCSPITQDGRVGS 120
Db 121 SAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
QY 121 TAVILDDNFVTKATALYDVPVYSSRHTIPQPFYSHSRYFTPKPVLDSIDYFQPNKR 180
Db 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLN 233
QY 181 NQLWRLQTSNVDHVGGLGTAFAENSKYDQDYNIRVTMYVQFREFNLKDPPLP 233

RESULT 5

ID O90239 PRELIMINARY; PRT; 230 AA.
AC O90239;

```

DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE CAPSID PROTEIN.
GN C1.
OS PORCINE CIRCOVIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF071879; G3551845;
SQ SEQUENCE 230 AA; 27494 MW; 09F65071 CRC32;

Query Match 66.3%; Score 1196; DB 14; Length 230;
Best Local Similarity 66.5%; Pred. No. 1.65e-228;
Matches 153; Conservative 31; Mismatches 43; Indels 3; Gaps 2;

Db 1 MTWPRRRRRRPRSHGLNLRPRRLAHPAFNRVRRKTKGIFNSRLSTELVLAIK 60
Qy 1 MTYPRRRRRRPRSHGLQLRRRPLVHP--RHRVRRKNGIFNTRLSRTFGYTVK 58
Db 61 -GGYSQSPWNVYKFNIGQFPLPPSGGNTPLPFPQYIRKAKYEFYPRDPTISNQRGV 119
Qy 59 RTVTTPSWAVDMFEKIDDFVPPGGGKNKISPEYIRKVKVFWPCSPITQDGRV 118
Db 120 GSTVVILDANFVPTSNLAYDPYINYSRHTIRQFTVHSRYFTPKPELDQIDWFHPNN 179
Qy 119 GSTAVILDNFVTATALTYPYVNVSSRHTIPQFVSYSRYFTPKPVLDSIDYFQPN 178
Db 180 KRNQLWHLNTHVNEHGLGVALONATQAQNVVRLIYVQFREFILKD 229
Qy 179 KRNQLWLRQTSNGVDHVLGTAFENSKYDQDYNIRVTMYVQFREFNLKD 228

RESULT 6 PRELIMINARY; PRT; 244 AA.
ID O90237
AC O90237;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
DE CAPSID PROTEIN (FRAGMENT).
GN C1.
OS BEAK AND FEATHER DISEASE VIRUS.
OC VIRUSES; SSNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RT "Beak and feather disease virus and porcine circovirus genomes:
intermediates between the geminiviruses and plant circoviruses."
RL ARCH. VIROL. 143:1723-1744(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA NIAGRO F.D., FORSTHOEFFEL A.N., LAWTHOR R.P., KAMALANATHAN L.,
RA RITCHIE B.W., LATIMER K.S., LUKERT P.D.;
RL SUBMITTED (JUN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AF071878; G3551842;
FT NON_TER 1
SQ SEQUENCE 244 AA; 28685 MW; 0164680E CRC32;

Query Match 12.9%; Score 233; DB 14; Length 244;
Best Local Similarity 31.7%; Pred. No. 2.59e-23;

```

```

atches. 77; Conservative 49; Mismatches 90; Indels 27; Gaps 21;

Db 1 RRRYARPYRRR-HI-RRYRRRR--HFRRR-RFS-TNRVYTLRLTRQFKFQIKQTST 64
Qy 5 RRRYRRRRPRSHGLQLRRRPLVHPRHRVRRKNGIFNTRLSRTFGYTV-KRTT 63
Db 65 GNLIFSSDVFVTFALSDFIT--NTPNPOTLNFENYRIKLAKMEMRPTWGHYTIQGD-GFGH 121
Qy 64 TP-SWAVDMFEKIDDFVPPGGGKNKISPEYIRKVKVFWPC-SPIT-QGDRGVGS 120
Db 122 TAVIODSRISKFTVADQTDQPLAPDGARKVTVSRGEK-RLLRPKPQITINDLTANQS 180
Qy 121 TAVIIDDNEVT-KATA-LTYDPIVNVSSRHTIPQFVSYSRHTIPRP-V-LDSTIDYFQP 176
Db 181 AALWLNARSARTGWIPLQGGPNAAGAKVRRHYGIGFSFPQPEQTITVYTKLIYVQFROPAPN 240
Qy 177 NKR-N-QL-WLRQ---TSG-NVDHVLGTAFENSKYDQDYNIRVTMYVQFREFNLK 227
Db 241 NPS 243
Qy 228 DPP 230

RESULT 7 PRELIMINARY; PRT; 770 AA.
ID O70800
AC O70800;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DT 01-AUG-1998 (TREMBLREL. 07, LAST ANNOTATION UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSNA VIRUSES; UNCLASSIFIED SSNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-G104901;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB011489; D1029096;
SQ SEQUENCE 770 AA; 90135 MW; ECE5EEB7 CRC32;

Query Match 8.1%; Score 146; DB 14; Length 770;
Best Local Similarity 32.1%; Pred. No. 6.62e-08;
Matches 26; Conservative 17; Mismatches 34; Indels 4; Gaps 4;

Db 10 RRRWRRRPRRRRRTTTRRRRRPARRRRNRVRRRGRRRRYRRWRKRRKRRKKAK 69
Qy 5 RRRYRRRRPRSHGLQLRRRPLVHPRHRVRRKNGIFNTRLSR-TF-GYTVKRTT 61
Db 70 IIRQWQPNYRRRCNIVGYP 90
Qy 62 VTTPSWAVD-WMRFKIDDFVP 81

RESULT 8 PRELIMINARY; PRT; 770 AA.
ID O70796
AC O70796;
DT 01-AUG-1998 (TREMBLREL. 07, CREATED)
DT 01-AUG-1998 (TREMBLREL. 07, LAST SEQUENCE UPDATE)
DE LONG ORF.
OS TT VIRUS.
OC VIRUSES; SSNA VIRUSES; UNCLASSIFIED SSNA VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-G103301;
RA TAKAHASHI K., OHTA Y., MISHIRO S.;
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AB011487; D1029092;
SQ SEQUENCE 770 AA; 90190 MW; A17FAS9 CRC32;

Query Match 7.9%; Score 143; DB 14; Length 770;
Best Local Similarity 32.1%; Pred. No. 2.03e-07;
Matches 26; Conservative 16; Mismatches 35; Indels 4; Gaps 4;

```


OM of: US-09-209-961-20 to: GenEmbl:* out_format : pfs

Date: Dec 27, 1999 2:13 AM

About: Results were produced by the GenCore software, version 4.5, Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

```

-Model=frame+g2n.model1 -DEV=xl1p
-o=/cgn2.1/USPRO_spool/US0929961/runat_22121999_101043_17910/app_query.fasta.1
-DB=GenEmbl1 -QMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000 -FGAPOP=4.500
-QCAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -START=1 -XGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTPMF=pfs -NORM=stat -USER=US0929961 -NCPU=6 -ICPU=3 -NO_XLPXY
-WAIT -THREADS=1

```

Search information block:

Query: US-09-209-961-20

Query: 03-03-203
Query length: 59

Query Length: 39
Database: GenEmbl:*

Database: GENEMOT:780561

Database sequences: 760361
Database length: 2137953050

Database length: 213793050
Search time (sec): 1994.760000

score list:

Sequence	Srdr	Orig	ZScore	BScore	Len	Documentation
gb_v1:AF085695	+	318.00	767.68	9.3e-35	1768	! AF085695 Porcine circovirus st
gb_v1:AF086834	+	318.00	767.68	9.3e-35	1768	! AF086834 Porcine circovirus st
gb_v1:AF086835	+	318.00	767.68	9.3e-35	1768	! AF086835 Porcine circovirus st
gb_v1:AF086836	+	318.00	767.68	9.3e-35	1768	! AF086836 Porcine circovirus st
gb_v1:AF027217	+	314.00	757.72	3.3e-34	1768	! AF027217 Porcine circovirus st
gb_v1:AF055391	+	314.00	757.72	3.3e-34	1768	! AF055391 Porcine circovirus st
gb_v1:AF055392	+	314.00	757.72	3.3e-34	1768	! AF055392 Porcine circovirus st
gb_v1:AF109397	+	314.00	757.72	3.3e-34	1768	! AF109397 Bovine circovirus, c
gb_v1:AF109399	+	314.00	757.72	3.3e-34	1768	! AF109399 Porcine circovirus, c
gb_v1:AF147751	+	314.00	757.72	3.3e-34	1768	! AF147751 Porcine circovirus ty
gb_v1:PCAJ3185	+	314.00	757.72	3.3e-34	1768	! AJ223185 Porcine circovirus DN
gb_v1:AF117753	+	310.00	747.75	1.2e-33	1768	! AF117753 Porcine circovirus ty
gb_v1:AF156729	+	310.00	747.75	1.2e-33	1768	! AF156729 Porcine circovirus, c
gb_v1:AF156528	+	310.00	747.75	1.2e-33	1768	! AF156528 Porcine circovirus, c
gb_v1:AF055393	+	307.00	740.28	3.1e-33	1767	! AF055393 Porcine circovirus ty
gb_v1:AF055394	+	307.00	740.28	3.1e-33	1767	! AF055394 Porcine circovirus ty
gb_v1:AF109398	+	306.00	737.79	4.3e-33	1768	! AF109398 Porcine circovirus ty
gb_v1:AF112862	+	305.00	735.29	5.9e-33	1768	! AF112862 Porcine circovirus ty
gb_v1:AF118095	+	292.00	712.17	1.1e-31	597	! AF118095 Porcine circovirus typ
gb_v1:AF071879	+	115.00	261.98	1.4e-06	1758	! AF071879 Porcine circovirus, c
gb_v1:PCCOMGEN	+	115.00	261.98	1.4e-06	1759	! Y09921 Porcine circovirus comp
gb_v1:AF012107	+	114.00	259.49	1.9e-06	1759	! AF012107 Porcine circovirus, c
gb_v1:PCU49186	+	114.00	259.49	1.9e-06	1759	! U49186 Porcine circovirus, c
gb_sts:HS064253	+	67.00	160.45	0.6171	212	! U64253 Human STS sequence NOTI-
gb_ba2:AF1533769	+	68.00	145.32	4.30	1671	! AF153769 Porphyromonas gingiva
gb_sts:HS1392204	+	60.00	138.73	10.00	350	! Z16877 H. sapiens (D7S505) DNA
gb_sts:YL1423296	-	58.00	138.32	10.54	237	! AF54381 Trifolium alexandrinum
gb_p12:AF154381	-	64.00	136.71	12.96	1426	! AJ242396 Yarrowia lipolytica S
gb_com:BTMFEN9	+	57.00	134.95	16.24	227	! Z83213 B.taurus microsatellite
gb_sts:CG09785	+	60.00	134.09	18.14	603	! G09785 human STS CHLC.GCT1005.
gb_ov:CHKLECN1	+	56.50	133.38	19.68	236	! Z22288 T.stenopetala DNA for in
gb_p11:TSNTS1	+	58.00	132.72	21.83	395	! M63230 Chicken hepatic lectin g
gb_ov:CHKLECN6	+	63.50	132.49	22.28	2021	! M4495 D.melanogaster Antennap
gb_inl1:DRONTPE4	-	63.50	132.36	22.65	2052	! X03790 D.melanogaster Antp gen
gb_inl1:DMANTPE4	+	61.50	132.24	22.99	1160	! X81542 Drosophila Antp recep
gb_pr3:S81542	+	61.50	132.24	22.99	1160	! X81542 Drosophila Antp recep
gb_sts:G50347	-	58.00	131.72	24.59	444	! G50347 SHGC-77761 Human Homo
gb_p12:AF154359	-	55.50	130.85	27.49	237	! AF154359 Trifolium hirtum inter
gb_ba2:PMU46125	+	58.00	130.44	28.98	516	! U46125 Pseudomonas mevalonii he
gb_sts:BTU78862	+	57.00	130.36	29.28	389	! U78862 Bos taurus transthyretin
gb_ro:086352	+	59.50	129.68	31.95	874	! D86352 Mouse mRNA for CPP32, c
gb_p12:AF154385	+	55.00	129.64	32.10	236	! AF154385 Trifolium pallidum int
gb_ov:AF182090	+	55.50	129.01	34.81	294	! AF182090 Scophthalmus maximus

/db_xref="GI:3668365"
/translation="MVLLNCCPSCRSSLSSEDFVLGILECYKTIHGGRGPVRHPPPP
P"

CDS

1016..1177
/codon_start=1
/product="P6.2"
/protein_id="AAC61863.1"
/db_xref="GI:3668366"
/db_xref="PID:3668366"
/translation="MVFIHLGFKWGVFKIKSELYIHGYTDIVLVVFTVPSAEAE
YVHHISRG"

CDS

complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61864.1"
/db_xref="GI:3668367"
/db_xref="PID:3668367"
/translation="MTYPRRRYRRHRRPSHLGQILRRPWLVPVHPRHRYRRKNGI
FNRLSRFTGYTVKATVTTPSWAVDMRFKIDFVPPGGTNGKISIPFVYRIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDPVNYSRHTIPQPSYHSRY
ETPKPVLDSIDYFQPNKRNLQWLRLQTSNVDHVLGTAFENSKYDODYINIRVTMY
VQREFNLKDPLEP"

BASE COUNT
ORIGIN

463 a 362 c 481 g 462 t

alignment_scores:

Quality: 318.00 Length: 59
Ratio: 5.390 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-20 x AF085695

Align seg 1/1 to: AF085695 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACTTCATTTGGGGCCACCTGGGTGGTGAAGCAATGGGCT 602

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAAAACCGGAAACACACATCTGGAACACCACTAAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAAAGTGGTTGTTATTGATGACTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59

703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF086834

seq_documentation_block:

LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.

ACCESSION AF086834

NID 93661515

VERSION AF086834.1 GI:3661515

SOURCE porcine circovirus.

ORGANISM porcine circovirus

DEFINITION Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES
source

1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
51..995
/codon_start=1

CDS

/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:93661516"
/db_xref="GI:3661516"
/translation="MPSKNGRSQPQHKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPHLQGFANFVKQTFNKVILGARCHIEKAGTDOQKEYSCKEGL
LLIECGAPRSQORSLSLSTAVSTLLESGLVTVAKOHPVTFKNFRGLAEILLKYSKGM
OKREDWKNYHFIVGPPCGCKSKWAANFETTYWPKPKNWDGIGHGEKVVYIDDFY
GMLPDDLLRLCDRYELTIVTKGTGTFVLARSILITSNQTPLEWISSSTAVFAVEALIR
RITSLVFNKNTQSTEEGGQFVTLSPPCPEFFYEIN"

CDS

complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:93661517"
/db_xref="GI:3661517"
/translation="MTYPRRRYRRHRRPSHLGQILRRPWLVPVHPRHRYRRKNGI
FNRLSRFTGYTVKATVTTPSWAVDMRFNIDFVPPGGTNGKISIPFVYRIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDPVNYSRHTIPQPSYHSRY
ETPKPVLDSIDYFQPNKRNLQWLRLQTSNVDHVLGTAFENSKYDQDYNIRVTMY
VQREFNLKDPPLKP"

BASE COUNT 463 a 357 c 482 g 466 t
ORIGIN

alignment_scores:

Quality: 318.00 Length: 59
Ratio: 5.390 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-20 x AF086834

Align seg 1/1 to: AF086834 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACTTCATTTGGGGCCACCTGGGTGGTGAAGCAATGGGCT 602

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAAAACCGGAAACACACATCTGGAACACCACTAAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAAAGTGGTTGTTATTGATGACTTT 702

51 MetAlaGlyCysArgGlyMetIleTyr 59

703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF086835

seq_documentation_block:

LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.

ACCESSION AF086835

NID 93661518

VERSION AF086835.1 GI:3661518

KEYWORDS porcine circovirus.

SOURCE porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES Location/Qualifiers
source
1. .1768
/organism="porcine circovirus"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="P35.8"
/protein_id="AAC61739.1"
/db_xref="PID:g3661519"
/db_xref="GI:3661519"
/translation="MPSKNGRSGPQPHKRWVTLNPNSENKKRIELPISLFDYFI
VGEENEGRTPLHGFANFVKQTFNKVYFGARCHIEKAKGTQDNKEYCSKEGN
LLIECGAPRSGQSDLSSTAVSLLESGLVTVAKQHPVTVFNKRLAELLKVSCKM
OKRDKNTNHFIVGPGCGSKWAANFANPETTYKPKKKNWDGYHGEKVVVDYFY
GWLPLDRLCLDRYPLTVKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATQSTEEGQFVTLSPPCPEFFEYIN"
complement(1034. .1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61740.1"
/db_xref="PID:g3661520"
/db_xref="GI:3661520"
/translation="MTYPRRRYRRRRHPRSHLQILRRRPLVHPHRYRWRRKNGI
FNRLSRTFGYTVKRTVTTPSWAVDMRFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCSPITGDRGVSTAVILDDNFVTKATALTYPYNYSSRHTIPQPSYHSRY
ETPKPVLDSIDYFOQNNKQLWLRLQTSQNVHDVGLGTAFENSKYDQYINIRVTY
VQREFNKLDPPEP"
BASE COUNT 464 a 360 c 480 g 464 t
ORIGIN

alignment_scores:
Quality: 318.00 Length: 59
Ratio: 5.390 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-20 x AF086835 ..
Align seg 1/1 to: AF086835 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACTTCATTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 602
|||||
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTGCAACCCGGAACCACTACTGGAACCACTTAAACAA 652
|||||
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTGTATTGATGACTTTT 702
|||||
51 MetaIaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729
|||||
seq_name: gb_vi:AF086836

seq_documentation_block:
LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain M226, complete genome.
ACCESSION AF086836
NID G3661521
VERSION AF086836.1 GI:3661521
KEYWORDS

SOURCE porcine circovirus.
ORGANISM porcine circovirus
REFERENCE 1 (bases 1 to 1768)
AUTHORS Viruses: ssDNA viruses; Circoviridae; Circovirus.
TITLE Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
JOURNAL Emergence of a new porcine circovirus
Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES Location/Qualifiers
source
1. .1768
/organism="porcine circovirus"
/strain="M226"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="P35.8"
/protein_id="AAC61741.1"
/db_xref="PID:g3661522"
/db_xref="GI:3661522"
/translation="MPSKNGRSGPQPHKRWVTLNPNSENKKRIELPISLFDYFI
VGEENEGRTPLHGFANFVKQTFNKVYFGARCHIEKAKGTQDNKEYCSKEGN
LLIECGAPRSGQSDLSSTAVSLLESGLVTVAKQHPVTVFNKRLAELLKVSCKM
OKRDKNTNHFIVGPGCGSKWAANFANPETTYKPKKKNWDGYHGEKVVVDYFY
GWLPLDRLCLDRYPLTVKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATQSTEEGQFVTLSPPCPEFFEYIN"
complement(1034. .1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61742.1"
/db_xref="PID:g3661523"
/db_xref="GI:3661523"
/translation="MTYPRRRYRRRRHPRSHLQILRRRPLVHPHRYRWRRKNGI
FNRLSRTFGYTVKRTVTTPSWAVDMRFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCSPITGDRGVSTAVILDDNFVTKATALTYPYNYSSRHTIPQPSYHSRY
ETPKPVLDSIDYFOQNNKQLWLRLQTSQNVHDVGLGTAFENSKYDQYINIRVTY
VQREFNKLDPPEP"
BASE COUNT 463 a 363 c 480 g 462 t
ORIGIN

alignment_scores:
Quality: 318.00 Length: 59
Ratio: 5.390 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-20 x AF086836 ..
Align seg 1/1 to: AF086836 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACTTCATTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 602
|||||
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTGCAACCCGGAACCACTACTGGAACCACTTAAACAA 652
|||||
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTGTATTGATGACTTTT 702
|||||
51 MetaIaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729
|||||
seq_name: gb_vi:AF027217
seq_documentation_block:

```

LOCUS       AF027217      1768 bp      DNA      circular      VRL      14-MAY-1998
DEFINITION  Porcine circovirus strain pmws PCV, complete genome.
ACCESSION   AF027217
NID         92689645
VERSION     AF027217.1  GI:2689645
SOURCE      porcine circovirus.
ORGANISM    Viruses: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE   1 (bases 1 to 1768)
AUTHORS     Hamel,A.L., Lin,L.L. and Nayar,G.P.
TITLE       Nucleotide sequence of porcine circovirus associated with
           postweaning multisystemic wasting syndrome in pigs
JOURNAL     J. Virol. 72 (6), 5262-5267 (1998)
MEDLINE     98241772
REFERENCE   2 (bases 1 to 1768)
AUTHORS     Hamel,A.L., Lin,L.L. and Nayar,G.P.S.
TITLE       Direct Submission
JOURNAL     Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services
           Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,
           Manitoba R3T 5S6, Canada
FEATURES    Location/Qualifiers
             1..1768
             /organism="porcine circovirus"
             /strain="pmws PCV"
             /db_xref="taxon:46221"
             /note="both strands of seven overlapping PCR fragments
             were sequenced: virus isolated from lung, lymph node,
             spleen and tonsil tissue from pigs affected by post
             weaning multisystemic wasting syndrome"
             51..995
             /note="ORF1; similar to Rep protein encoded by
             non-pathogenic PCV, GenBank Accession Number U49186;
             predicted 35.8 kDa protein"
             /codon_start=1
             /product="putative Rep protein"
             /protein_id="AAC59462.1"
             /db_xref="PID:g2689646"
             /db_xref="GI:2689646"
             /translation="MPSKNGRSGQPQKRWVFTLNNSDEKRIKRLPISLDFYFI
             VGEENEGRTPLHGFANVRKQTFNKVYLGARCIYKAKGIDQNKKEYCSKEGN
             LLIEGAPRSQSDLTAVTLLESGSLTVAEQHPVTYRNFRLGLAELLKVGKM
             QKRDKNVHVIVGPGGCKSKAAAFADPTTYWPKRWMDGYHGEVVVIDFY
             GWLPWDLRLCDRLPLTVETKGGTVPLFARSLITSNOTPLEWYSSTAVPAVALYR
             RTSLVFNKNATEQSTEGGQFTLSPPCPPEYIN"
             117..125
             /note="glycosylation site"
             complement(357..671)
             /note="ORF3; predicted 11.9 kDa protein"
             /codon_start=1
             /protein_id="AAC59464.1"
             /db_xref="PID:g2689648"
             /db_xref="GI:2689648"
             /translation="MVTIPPLVSRWFPGVGRVCKISSPFAFTTPRWPNHNDVYIGLPI
             TLLHPFAHFQKESQPAEISDKRYVLLCNGHOTPALQOQTHSSRQVTPLSRSSTF
             NK"
             complement(386..565)
             /note="ORF4; predicted 6.5 kDa protein"
             /codon_start=1
             /protein_id="AAC59465.1"
             /db_xref="PID:g2689649"
             /db_xref="GI:2689649"
             /translation="MTCILVFQSRFCIFPLTFKSSAPRKFLTNVTGCSATVTRLPL
             SNKVLTAVDRSLRCP"
             complement(470..478)
             /note="glycosylation site"
             complement(688..753)
             /note="ORF8; predicted 2.3 kDa protein"
             /codon_start=1
             /protein_id="AAC59469.1"
             /db_xref="PID:g2689653"
             /db_xref="GI:2689653"

```

```

             /translation="MDIDHTVSDHPTAAASHKSHQ"
             816..824
             /note="glycosylation site"
             906..914
             /note="glycosylation site"
             983..988
             complement(989..1033)
             /note="ORF1; predicted 1.8 kDa protein"
             /codon_start=1
             /protein_id="AAC59472.1"
             /db_xref="PID:g2689656"
             /db_xref="GI:2689656"
             /translation="MKNNHYEVIKKTQ"
             1016..1177
             /note="ORF5; predicted 6.2 kDa protein"
             /codon_start=1
             /protein_id="AAC59466.1"
             /db_xref="PID:g2689650"
             /db_xref="GI:2689650"
             /translation="MVFIFHLGFKMGVFKIKFSELYIHGYTDIIVLVVTVTFERSAEA
             YVHISGRG"
             complement(1022..1027)
             complement(1034..1735)
             /note="ORF2; predicted 27.8 kDa protein"
             /codon_start=1
             /protein_id="AAC59463.1"
             /db_xref="PID:g2689647"
             /db_xref="GI:2689647"
             /translation="MTYPRRYRRRRPSPHGLQILRRPWLPHPRHYRWRKNGKI
             FNRJLSRTFGYTKATVTPSWAVDMRFNIDDFVPGGNTKISIPFYIRKRYK
             VEFWPCSPITQDGRGVSTAVILDDNFVKATALTDYDVYVNSRHTIPQPFYHSRY
             FTKPVLDSTIDYFQPNKRTQLWLRLQTSRNVHDVHVLGTAFAENSIYDQDYNIRVTMY
             VQFREFNLKDPPLKP"
             complement(1301..1309)
             /note="glycosylation site"
             complement(1522..1611)
             /note="ORF6; predicted 3.1 kDa protein"
             /codon_start=1
             /protein_id="AAC59467.1"
             /db_xref="PID:g2689651"
             /db_xref="GI:2689651"
             /translation="MASTPASPAPSDILSRLPQSRPPGRWT"
             1524..1631
             /note="ORF10; predicted 4.1 kDa protein"
             /codon_start=1
             /protein_id="AAC59471.1"
             /db_xref="PID:g2689655"
             /db_xref="GI:2689655"
             /translation="MSTAQEGVLTVVALTVPVKVRRRLKMPFFLIQR"
             1682..1741
             /note="ORF7; predicted 1.9 kDa protein"
             /codon_start=1
             /protein_id="AAC59468.1"
             /db_xref="PID:g2689652"
             /db_xref="GI:2689652"
             /translation="MAAGAVSSSAVTPPWIRHS"
             complement(1732..1768,1..92))
             /note="ORF9; predicted 4.6 kDa protein"
             /codon_start=1
             /protein_id="AAC59470.1"
             /db_xref="PID:g2689654"
             /db_xref="GI:2689654"
             /translation="MWLGSSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRQL"
             join(1746..1768,1..13)
             /function="putative replication site"
             join(1762..1768,1..2)
             /note="'AAGTATTAC' is similar to the nonanucleotide motif
             in the non-pathogenic PCV, GenBank Accession Number
             U49186"

```

BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:

Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:

US-09-209-961-20 x AF027217 ..
Align seg 1/1 to: AF027217 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACAGCTCATGTGGGGCCACCTGGGTGTGTAAGAAAGCAATGGCGT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTCAGACCCGGAACACACATACATGGAACACCACTAGAACAA 652
34 erGlyGlyMetValThrMetValLysLysTyrLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGATGTTACCATGCTGAAGAGTGGTGTCTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF055391

seq_documentation_block:

LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from USA, complete genome.
ACCESSION AF055391
NID 93598796
VERSION AF055391.1 GI:3598796
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
porcine circovirus
viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A., Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting syndromes in pigs
J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
JOURNAL
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The Queen's University of, Belfast, Stoney Road, Belfast, Northern Ireland, BT4 3SD, U.K.
FEATURES
source
1..1768
Location/Qualifiers
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.999 EcorI No.8"
/country="USA"
complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:g3598807"
/db_xref="GI:3598807"
/translation="MASTFASPAPSDILSRLPQSERPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:g3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVALTYPKVRRRLKMPFFLLQR"

CDS

261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"
/db_xref="PID:g3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(311..439)
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGSSASSILLAGHVAEVLPRCCRCRSALVILTAHFRFQL"

CDS

398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKKNRSGPQPHKRVFTLNPFSEDERKKIRELPISLFDYFI
VGEENEGRTPHLQGFANFYKQTFNKWKYLGARCHIEKAKGTDQONKEYCSKEGN
LLIEGAPRSQGRSDLSSTAVSTLLESGSLVVAEQHPVTFYRNRGLAEILKVGKM
OKRDKTNVRVIVGPPGCGSKWAANFADPTTYWKPKNKWDGYPHGEVYIDDFY
GWLPMDDLRLCDRYPLTIVETKGTVPFLARSILITISNOTPLEWYSSTAVPAVEALYR
RITSLVFNKNAEQSTEEGQFVTLSPPCPEFFEYNY"
674..679
complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTTPRPHNDVYIGLPI
TLHFFPAHFKFSQPAEISDKRYRLLCNGHOTPALQOQTHSSROVTPILSRSSIF
NK"

polyA_signal
CDS

complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCTLVQSRFCIFPLTFKSSASPRKFLTNVTGCCATVTRLPL
SNKVLTAVDRLRCP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDVDPHTAAASHKHQ"

CDS

1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKKQT"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVKIKFSELYINGYTDIVVLVYVTVFERSAEA
YVWHISRGL"

CDS

complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDVDPHTAAASHKHQ"

polyA_signal
CDS

complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKKQT"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVKIKFSELYINGYTDIVVLVYVTVFERSAEA
YVWHISRGL"

CDS

complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDVDPHTAAASHKHQ"

```
polyA_site
CDS
complement(1369..1374)
complement(join(1381..1768,1..314))
/clone="Imp.1010-Stoon ECORI No.14"
/country="Canada"
complement(101..190)
/codon_start=1
/product="ORF2"
/protein_id="AAC35299.1"
/db_xref="PID:g3598798"
/db_xref="GI:3598798"
/translation="MTYPRRRYRRRRHPRSHLGLRRRPLVHPRHRYRWRKNGI
ENTLSRFTGTVKATVTTPSWAVDMRMENIDDFVPGGCTNKISIPPEYRIKKVK
VEFPWCSPTGQDRCVGSFVILDDNFYTKATALYDYPVYSSRHTIQPQFSYHSRY
ETPRVLDSTIDIFQPNKRNQMLURLTNRVHVLGTFAPESIIYDQYNIIRVTM
VOFRENLKDPPLKP"
BASE COUNT 453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:
  Quality: 314.00      Length: 59
  Ratio: 5.322        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x AF055391 ..
Align seg 1/1 to: AF055391 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValVallysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATGTGGGCCACCTGGTGTGTAACAAACAAATGGGCT 949
|||||
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrs 34
|||||
950 GCTAATTTTCACACCGGAAACACATACATCGGAACACCTAGAACAA 999
|||||
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
1000 GTGCTGGGATGTTACATGCTGAAGAAGTGTGTATTGATGACTTTT 1049
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGCTGGCTGCGCGCGGTGGATGATCTAC 1076

seq_name: gb_vi:AF055392

seq_documentation_block: 1768 bp DNA circular VRL 13-SEP-1998
LOCUS AF055392
DEFINITION Porcine circovirus Type II from Canada, complete genome.
ACCESSION AF055392
NID g3598808
VERSION AF055392.1 GI:3598808
KEYWORDS
SOURCE porcine circovirus.
ORGANISM
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
  Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
  Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
  Characterization of novel circovirus DNAs associated with wasting
  syndromes in pigs
  J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
JOURNAL 98418498
MEDLINE
REFERENCE 2 (bases 1 to 1768)
  Meehan,B.M.
  Direct Submission
  Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
  Queen's University of, Belfast, Stoney Road, Belfast, Northern
  Ireland, BT4 3SD, U.K.
  Location/Qualifiers
  1..1768
  /organism="porcine circovirus"
  /specific_host="Sus scrofa"
  /strain="Porcine circovirus Type II"
```

```

CDS      complement(1336..1380)
        /note="predicted 1.8 kDa protein"
        /codon_start=1
        /product="ORF1"
        /protein_id="AAC35316.1"
        /db_xref="PID:g3598816"
        /db_xref="GI:3598816"
        /translation="MNNKNHYEVKKTKQ"
        1363..1524
CDS      /note="predicted 6.2 kDa protein"
        /codon_start=1
        /product="ORF5"
        /protein_id="AAC35313.1"
        /db_xref="PID:g3598813"
        /db_xref="GI:3598813"
        /translation="MVFIIHLGFRWGVKIFSELYIHGYTDIVVLVVTTFERSAEA
        YVHISRGL"
        complement(1369..1374)
        /note="predicted 27.8 kDa protein"
        /codon_start=1
        /product="ORF2"
        /protein_id="AAC35310.1"
        /db_xref="PID:g3598810"
        /db_xref="GI:3598810"
        /translation="MTVPRRYRRRRHRRPRSHLQILRRPWLVPVHPRHYRWRKNGI
        FNTLSFTGTVKRTVTPSWAVDMRFKIDDFVPGGTNKISIFPEYVRIKVK
        FFWPCSPITQDGVGSGTAVILDNFTVKATALYDPVINYSSRHTIPQFSTHSRY
        VFKPVLVDSTIDYFQPNKRNQNLWLRLQTSNDVRVGLGAFAFENSKYDQDYNIRVTMY
        VQREFNLKDPPLKP"
        451 a 361 c 495 g 461 t

BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 314.00      Length: 59
  Ratio: 5.322        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 98.305

alignment_block:
  US-09-209-961-20 x AF055392 ..

  Align seg 1/1 to: AF055392 from: 1 to: 1768

    1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaasnGlyLe 17
      |||||
    900 ATGTACACGTCATGTGGGCCACCTGGGTGTGTAATAAGCAATGGCT 949
      |||||

    17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
      |||||
    950 GCTAATTTGACACCGGACCCGACCATACATAGTGAACACCTAGAACAA 999
      |||||

    34 exGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
      |||||
    1000 GTGGTGGGATGTTACCATGGGTGAGAGAGTGTGTATGATGACTTTT 1049
      |||||

    51 MetAlaGlyCysArgGlyMetIleTyr 59
      |||||
    1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076
      |||||

seq_name: gb_vi:AF109397

seq_documentation_block:
  LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
  DEFINITION Bovine circovirus, complete genome.
  ACCESSION AF109397
  NID 94210313
  VERSION AF109397.1 GI:4210313
  KEYWORDS
  ORGANISM bovine circovirus.
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 1768)

```

Hamel,A.L. and Nayar,G.P.S.
 Nucleotide sequence of a circovirus detected in cattle with various
 clinical syndromes
 Unpublished
 2 (bases 1 to 1768)
 Hamel,A.L. and Nayar,G.P.S.
 Direct Submission
 Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
 Laboratory, Manitoba Agriculture, 545 University Crescent,
 Winnipeg, Manitoba R3T 5S6, Canada
 3 (bases 1 to 1768)
 Hamel,A.L. and Nayar,G.P.S.
 Direct Submission
 Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
 Laboratory, Manitoba Agriculture, 545 University Crescent,
 Winnipeg, Manitoba R3T 5S6, Canada
 Sequence update by submitter
 On Feb 2, 1999 this sequence version replaced gi:4028609.
 Location/Qualifiers
 1..1768
 /organism="bovine circovirus"
 /viral
 /db_xref="taxon:85542"
 /note="BCV
 sequence was obtained from several overlapping PCRs using
 DNA extracted from various tissues from cattle; similar to
 Porcine circovirus sequence presented in GenBank Accession
 Number AF027217"
 13..35
 /rpt_type=tandem
 /rpt_unit=13..18
 51..995
 /note="ORF-1; similar to Porcine circovirus ORF1 encoded
 by the sequence presented in GenBank Accession Number
 AF027217."
 /codon_start=1
 /product="putative Rep and coat protein"
 /protein_id="AAD11928.1"
 /db_xref="PID:g4210314"
 /db_xref="GI:4210314"
 /translation="MPSKKNRSGPQPHKRWFTLNNPSEDEKKIKELPISLFDYFI
 VGEENEGRTPHLOGFANFKVKTQFVKVYLGARCHIEKAKGDOONKEYCSKGN
 LIIEGAPRSQGSQSDLSFVSTLLESGSLTVAEQHPVTFVFNFGELALLKVSCKM
 QKRDKNVHVIVGPPCGCKSKWANFADPETTYKPPKKNKWDGTHGEVWIDDFY
 GWLPWDLRLCNDRYPLTVETGKGVTFVFLARSILITSNOTPLEWISSTAVPAVEALYR
 RITSLVFNKNAEQSTEKGQFVTLSPPCPEPPYEINY"
 327..332
 complement(357..671)
 /note="similar to Porcine circovirus ORF-3 encoded by the
 sequence presented in GenBank Accession Number AF027217."
 /codon_start=1
 /product="ORF-3"
 /protein_id="AAD11930.1"
 /db_xref="PID:g4210315"
 /db_xref="GI:4210315"
 /translation="MVTIPLVSRMFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
 TLLHFPFHQFQSPAEISDKRYRLVLLNGHOTPALQOQTHSSRQVTPLSLRSSSTF
 NK"
 complement(386..565)
 /note="similar to Porcine circovirus ORF-4 encoded by the
 sequence presented in GenBank Accession Number AF027217."
 /codon_start=1
 /product="ORF-4"
 /protein_id="AAD11931.1"
 /db_xref="PID:g4210316"
 /db_xref="GI:4210316"
 /translation="MTCILVFSQRCIFPLTFKSSASPRKFLTNVTCGCSATVRLPL
 SNKVLTAVDRSLRCP"
 553..732
 /codon_start=1
 /product="ORF-12"
 /protein_id="AAD11927.1"
 /db_xref="PID:g4028610"

AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 REMARK
 COMMENT
 FEATURES
 source
 repeat_region
 CDS
 polyA_signal
 CDS
 CDS
 CDS


```

/db_xref="GI:4028610"
/translation="MTSLWGLGVVVKANGLLILOTRKPHGTGNHLETSGGVMVWKVK
LLIMTFMAGCRGMIV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTVSVDHPTAAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNNKHVEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:g4028615"
/translation="MVFIIHLGFKWGVFKIFSELYIHGYTDIVLVLVFTFERSAEA
YVVIHSRGL"
complement(1022..1027)
complement(1256..1735)
/notes="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:g4028612"
/db_xref="GI:4028612"
/translation="MTYPRRRYRRTRPRSHGILRRPWLPHRRYRNRKNGI
FNTLSRTGYTKATVTPTSWAVDMRFNIDDFPPGGTNKISIPFEYIRKVK
VFVPCSPITQDGRGVSTAVILDDNFKATALTVDYVYVNSRRTIPQPFYSYHR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASTPASPAPSDILSLRPLQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="MSTAEGVLTVVATVYPKRERVVKMPFFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:g4028617"
/translation="MAAGAGPSSAVTPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 452 a 362 c 493 g 461 t
ORIGIN

alignment_scores:
  Quality: 314.00
  Ratio: 5.322
  Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

```

```

alignment_block:
US-09-209-961-20 x AF109397
Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCAATCTGGGGCCACCTGGGTGTGGTAAAGCAATGGGCT 602
|||||
17 uleuileLeuGlnThrArgLysProHisThrClyAsnHisLeuLysThr 34
|||||
603 GCTAATTTTGCAGACCCGGAACACCATCTGGAAACCACTAGAAACAA 652
|||||
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
653 GTGGTGGATGGTTACCATGGTGAAGAAGTGGTCTTATTGATGACTTTT 702
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729
|||||

seq_name: gb_vi:AF109399

seq_documentation_block:
LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2-E.
porcine circovirus type 2-E
Viruses; ssDNA viruses; Circoviridae; Circovirus; #69#line
circovirus type 2
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-E"
/virion
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
1..995
/notes="ORF-1."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKNGRSGPQPKRWFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRTPHLQGFANVKKQTFNKVWYLCARCHIEKAKGTDOQNKKEYCSKEGN
LLIECAPRSQGRSDLSVTLSEGLVTVAEQHPVTFVRNFRGLAEILKVSGRM
OKRDKWTKNVHIVGPPCGSKWAANFADPTTYKPKRWKWDGIVGEEVYVDDFY
GILPDDLLRLCDRIPLTVETKGTGTFPLARSILITSPLEWTPSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFEYIN"
327..332
polyA_signal

```


/isolate="ISUVDL 98-15237"
/specific_host="Sus scrofa"
/db_xref="taxon:85708"
/note="Isolated from pigs with postweaning multisystemic
wasting syndrome"

CDS

51..995
/note="ORF1: 35-36kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="PID:G5007011"
/db_xref="GI:5007011"
/translation="MPSKNGRSGQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKVKTQFNKVLGARCHIEKAKGTDOONKEKCSKEGN
LLIEGAPSGORSGLSTAVSLLESGLVTAEOHPVTFVNRFLAELLKVSCKM
QRDKNTNVHVIVGPPCGCKSKAANFADPETTYWKPNNKWDGHHGEVVDYFI
GWLPMDDLRLCDRLVETKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFYEINY"
complement(1034..1735)
/note="ORF2: 27-28kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="PID:G5007012"
/db_xref="GI:5007012"
/translation="MTYPRRRYRRRRRPRSHLQILRRPMLVPHRRYRWRKNGI
FNTRLSRTGYVKAITVTTPSNVDMRFNIDDFVPPCGGNTKISIPETTYRIRVK
VFEPSPSTQDGRGVSTAVILDDNFNTAKATIDYDYNYSRRITIPQPSCHRY
FEPKPLVDSTIDYFQPNRRNQLWLRSLQTSRNVHDVGLGTAFAENSIYDQYNIRVTMY
VQFREFNLKDPPLKP"

BASE COUNT 450 a 362 c 494 g 462 t
ORIGIN

alignment_scores:

Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:

US-09-209-961-20 x AF147751 ..

Align seg 1/1 to: AF147751 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATTGTGGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTGACACCGGAAACCACTACTGGAACCACTAGAAACAA 652
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGTTACCATGTTGAAGAGTGGTGTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_v1:PCAJ3185

seq_documentation_block:

LOCUS PCAJ3185 1768 bp DNA circular VRL 03-JUL-1998
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION AJ223185
NID 93293040
VERSION AJ223185.1 GI:3293040
KEYWORDS complete genome.
SOURCE porcine circovirus.
ORGANISM porcine circovirus
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1. (bases 1 to 1768)

AUTHORS
TITLE
JOURNAL

Morozov, I.
Direct Submission
Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
50011, U.S.A

REFERENCE

AUTHORS

2 (bases 1 to 1768)
Morozov, I., Sirinarumitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
Yoon, K.J. and Paul, P.S.

TITLE

Detection of a novel strain of Porcine circovirus in pigs with

JOURNAL

Postweaning Multisystemic Wasting Syndrome (PMWS)

FEATURES

Location/Qualifiers

1..1768
/organism="porcine circovirus"
/virus
/isolate="ISU-31"
/specific_host="Sus scrofa"
/db_xref="taxon:46221"
51..995
/note="ORF1"
/codon_start=1
/protein_id="CAAL1157.1"
/db_xref="PID:e1310034"
/db_xref="GI:3293041"
/translation="MPSKNGRSGQPQHRKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKVKTQFNKVLGARCHIEKAKGTDOONKEKCSKEGN
LLIEGAPSGORSGLSTAVSLLESGLVTAEOHPVTFVNRFLAELLKVSCKM
QRDKNTNVHVIVGPPCGCKSKAANFADPETTYWKPNNKWDGHHGEVVDYFI
GWLPMDDLRLCDRLVETKGTGTFPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPPCPEFFYEINY"

CDS

452 a 361 c 492 g 463 t
BASE COUNT
ORIGIN

alignment_scores:

Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:

US-09-209-961-20 x PCAJ3185 ..

Align seg 1/1 to: PCAJ3185 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATTGTGGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTGACACCGGAAACCACTACTGGAACCACTAGAAACAA 652
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGTTACCATGTTGAAGAGTGGTGTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_v1:AF117753

seq_documentation_block:

LOCUS AF117753 1768 bp DNA circular VRL 04-FEB-1999
DEFINITION Porcine circovirus type 2-D, complete genome.
ACCESSION AF117753
NID 94219093
VERSION AF117753.1 GI:4219093
KEYWORDS
SOURCE porcine circovirus type 2-D.
ORGANISM porcine circovirus type 2-D.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus; #;#ine

```

circovirus type 2.
1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Genetic characterization of four novel type-2 Porcine circoviruses
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1999) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
Location/Qualifiers
1. .1768
/organism="porcine circovirus type 2-D"
/virion
/db_xref="taxon:86385"
/notes="type 2-D designation is based upon restriction
endonuclease digestion pattern
several overlapping PCR fragments were sequenced; virus
isolated from lung, mesenteric lymph node and tonsil
tissue; similar to Porcine circovirus sequence presented
in GenBank Accession Number AF027217"
repeat_region 13. .35
/rpt_type=tandem
/rpt_unit=13. .18
51. .995
/notes="putative Rep and coat protein"
/codon_start=1
/product="ORF-1"
/protein_id="AAD12308.1"
/db_xref="PID:g4219094"
/db_xref="GI:4219094"
/translation="MPSKNGRSGPOPKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPHLOGFANFYKQTEKNKYWLGACSHIEKAGTQONKEYCSKEGN
LMCEGAPRSQGSRLSTAVSTLSEGLSVTVAEOHPVTFVRNFGLAELLKVSGRM
KQKRWNTNHHVIVPGCCSKWAANFADPTTYPKPNKWDGHEVVVVIDDFY
GWLWDLLRLCDLTPLVETKGTVPFADPILTSITSNOTPLEWYSSAAVPAVEALYR
RITSLVFNKATEQSEGGQFVTLSPPCPEFFYEINY"
327. .332
complement(357. .671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD12309.1"
/db_xref="PID:g4219095"
/db_xref="GI:4219095"
/translation="MTIPLVSRWTFVCGFRVCKISSPPAFATPRWPHNDVYIRLPI
TLHLFPAHFQKFSQPAIEDKRYRVLLCNGHOTPAKQGTGHSGRQVTPLSRSRSTF
HQ"
complement(386. .565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD12310.1"
/db_xref="PID:g4219096"
/db_xref="GI:4219096"
/translation="MTCTFVQSRFCIFPLTFKSSASPRKFLNVTGCCSATVTRLPL
SSKVLTAVDRLRCP"
553. .732
/codon_start=1
/product="ORF-12"
/protein_id="AAD12311.1"
/db_xref="PID:g4219097"
/db_xref="GI:4219097"
/translation="MTSLWGHIGVAKANGLLILQTRKPHGTGHNHLETSGGMVTWVKW
LLMTFMAGCRGMIV"
complement(688. .753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD12312.1"
/db_xref="PID:g4219098"
/db_xref="GI:4219098"
/translation="MDTDHTVSDVHDPTAASHKSHQ"
983. .988
complement(989. .1033)
/codon_start=1

```

```

/polyA_signal
CDS
/product="ORF-11"
/protein_id="AAD12313.1"
/db_xref="PID:g4219099"
/db_xref="GI:4219099"
/translation="MNNKNHYEVIKKTO"
1016. .1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD12314.1"
/db_xref="PID:g4219100"
/db_xref="GI:4219100"
/translation="MVFIHLGFKWGVKIKFSELYIHGYTDIVVLVYTVFEERSAEA
YMYVSISSL"
complement(1022. .1027)
complement(1256. .1735)
/notes="similar to Porcine circovirus ORF-2 product encoded
by the sequence presented in GenBank Accession Number
AF027217"
/codon_start=1
/product="ORF-2"
/protein_id="AAD12315.1"
/db_xref="PID:g4219101"
/db_xref="GI:4219101"
/translation="MTFPFRRYRRRRHRRPSHLGQILRRRPWFHLPRHRYRWRKNGI
ENRLSRFTGYTKATVTPSWAVDMLRFKIDDFVPGGNTKISIPFEYVIRKVK
VEFWPCSPITQDGRGVGSSAVILDNFFPKSTALTYDPYVNYSSRHTITQPFYSHR"
1524. .1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD12317.1"
/db_xref="PID:g4219103"
/db_xref="GI:4219103"
/translation="MSTAQEGVLTVALTVTPKVRERVLMKPFELLOR"
complement(1528. .1611)
/product="ORF-6"
/protein_id="AAD12316.1"
/db_xref="PID:g4219102"
/db_xref="GI:4219102"
/translation="MASSTPASPAPSDILSRLOSERPPGR"
1682. .1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD12318.1"
/db_xref="PID:g4219104"
/db_xref="GI:4219104"
/translation="MAAGAVSSSAVTPPWKRHS"
join(1750. .1768.1. .13)
/notes="putative replication site"
join(1762. .1768.1. .2)
/notes="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186."
BASE COUNT 452 a 358 c 497 g 461 t
ORIGIN

```

```

alignment_scores:
Quality: 310.00 Length: 59
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.610
alignment_block:
US-09-209-961-20 x AF117753 ..
Align seg 1/1 to: AF117753 from: 1 to: 1768
1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATTGTGGGCCACCTGGCTGTGGCAAAAGCAATGGCT 602
|||||
17 UteuLeuLeuGlnThrArgGlyProHisThrGlyAsnHisLeuLysThrs 34
|||||

```

603 GCTAATTTTCAGACCCGGAACACATACACTGGAACACCACTAGAACAA 652
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTGTTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF154679

seq_documentation_block: 1768 bp DNA circular VRL 13-JUN-1999
LOCUS AF154679
DEFINITION Porcine circovirus, complete genome.
ACCESSION AF154679
NID 95052004
VERSION AF154679.1 GI:5052004
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Kuo,T.Y., Chou,Y.C. and Lai,S.S.
TITLE Complete nucleotide sequences analysis of porcine circovirus outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Kuo,T.Y., Chou,Y.C. and Lai,S.S.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan University, 142, Chousan Road, Taipei 106, Taiwan

FEATURES
source
1..1768
/organism="porcine circovirus"
/isolate="MLT98"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/note="putative Rep protein"
/codon_start=1
/product="ORF1"
/protein_id="AAD38398.1"
/db_xref="PID:g5052005"
/db_xref="GI:5052005"
/translation="MPSKNGRSGQPQHKRWFTLNPNSEDERKKIRELPISLDFCFI
VGEENEGRTPLQGFANFKQTFNKKVYFGARHIEKAGTDQONKEYCSKEGN
LLIECGAPRSQQRSDLSAVSTLSESGSLVTAQHPVTFVRNFRGLAELLKVSQKM
OKRDKNTNHHVIVGPPCGKSKWAANFADPTTYWKPPRNKWDYHGEEVVIDDFY
GWLWDLLRLCDRPLTVETKGTVPFLARSILITSNQTPLEWYSSSTAVPAVEALYR
RITSLVFNKNATEQSTEGGQFVTLSPCPPEFYINY"

BASE COUNT 451 a 367 c 495 g 455 t
ORIGIN

alignment_scores:
Quality: 310.00 Length: 59
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.610

alignment_block:

US-09-209-961-20 x AF154679 ..

Align seg 1/1 to: AF154679 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATCTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAGACCCGGAACACATACACTGGAACACCACTCGAACAA 652

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
653 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTGTTATTGATGACTTTT 702
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
703 ATGGCTGGCTGCCGTGGGATGATCTAC 729

seq_name: gb_vi:AF166528

seq_documentation_block: 1768 bp mRNA 19-AUG-1999
LOCUS AF166528
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID 95739338
VERSION AF166528.1 GI:5739338
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang,K.H., Lee,Y.F., Chao,D.S., Shieh,Y.C. and Lai,S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-Yi City 600, Taiwan

FEATURES
source
1..1768
/organism="porcine circovirus"
/viroion
/strain="Tainan"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/codon_start=1
/product="putative Rep protein"
/protein_id="AAD50432.1"
/db_xref="PID:g5739339"
/db_xref="GI:5739339"
/translation="MPSKNGRSGQPQHKRWFTLNPNSEDERKKIRELPISLDFCFI
VGEENEGRTPLQGFANFKQTFNKKVYFGARHIEKAGTDQONKEYCSKEGN
LLIECGAPRSQQRSDLSAVSTLSESGSLVTAQHPVTFVRNFRGLAELLKVSQKM
OKRDKNTNHHVIVGPPCGKSKWAANFADPTTYWKPPRNKWDYHGEEVVIDDFY
GWLWDLLRLCDRPLTVETKGTVPFLARSILITSNQTPLEWYSSSTAVPAVEALYR
RITSLVFNKNATEQSTEGGQFVTLSPCPPEFYINY"

BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN

alignment_scores:
Quality: 310.00 Length: 59
Ratio: 5.254 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 96.610

alignment_block:

US-09-209-961-20 x AF166528 ..

Align seg 1/1 to: AF166528 from: 1 to: 1768

1 MetTyrThrSerLeuTTPGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
553 ATGTACACGTCATCTGGGGCCACCTGGGTGGTAAAGCAATGGGCT 602
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
603 GCTAATTTTCAGACCCGGAACACATACACTGGAACACCACTCGAACAA 652
34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50

|||||
 653 GTGGTGGATGTTACCATGTTGAAGAGTGGTTGTTATTGATGACTTTT 702

51 MetalaGlyCysArgGlyMetIleTyr 59
 |||||
 703 ATGGTGGTGCCTCCGTTGGATGATCTAC 729

seq_name: gb_vi:AF055393

seq_documentation_block: 1767 bp DNA circular VRL 13-SEP-1998
 LOCUS AF055393 Porcine circovirus Type II from France, complete genome.
 DEFINITION Porcine circovirus Type II from France, complete genome.
 ACCESSION AF055393
 NID 93598820
 VERSION AF055393.1 GI:3598820
 KEYWORDS
 SOURCE
 ORGANISM
 Viruses; ssDNA viruses; Circoviridae; Circovirus.
 REFERENCE 1 (bases 1 to 1767)
 AUTHORS Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A., Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
 TITLE Characterization of novel circovirus DNAs associated with wasting syndromes in pigs
 JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
 MEDLINE 98418498
 REFERENCE 2 (bases 1 to 1767)
 AUTHORS Meehan,B.M.
 TITLE Direct Submission
 JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The Queen's University of, Belfast, Stoney Road, Belfast, Northern Ireland, BT4 3SD, U.K.

FEATURES

source
 1..1767
 /organism="porcine circovirus"
 /specific_host="Sus scrofa"
 /strain="Porcine circovirus Type II"
 /db_xref="taxon:46221"
 /clone="Imp.1011 48121 EcoRI No.17"
 /country="France"
 complement(101..190)
 /codon_start=1
 /product="ORF6"
 /protein_id="AAC35329.1"
 /db_xref="PID:g3598830"
 /db_xref="GI:3598830"
 /translation="MASSTPASPAPSDILSSPQSRPPGRT"
 103..210
 /codon_start=1
 /product="ORF10"
 /protein_id="AAC35327.1"
 /db_xref="PID:g3598828"
 /db_xref="GI:3598828"
 /translation="MSTAQEGVLTVVRLTVYKVRRLVKWPFLLQR"
 261..320
 /codon_start=1
 /product="ORF7"
 /protein_id="AAC35328.1"
 /db_xref="PID:g3598829"
 /db_xref="GI:3598829"
 /translation="MAAGAVSSPVTPPIRIH"
 complement(311..439)
 /codon_start=1
 /product="ORF9"
 /protein_id="AAC35325.1"
 /db_xref="PID:g3598826"
 /db_xref="GI:3598826"
 /translation="MGLGSASSILLARHVAEVLPRCCRSALVILTAHFFRFQI"
 398..1342
 /codon_start=1
 /product="putative Rep protein (ORF1)"
 /protein_id="AAC35320.1"

/db_xref="PID:g3598821"
 /db_xref="GI:3598821"
 /translation="MPSKNGSGPQPKRWFTLNNPSEDERKKIRDLPLSLDFEI
 VEGNEEGTPIHOGFANFVKQTFNKVWYLGARCHIERAKGTDOQNKYCSKEGN
 LMCEGAPRQSGQSDLSAVSTILLESGLTVAEQHPVIFVRNFRGLAELLKVSGRM
 OKRDWKTNVHIVGPPGCGSKWAANFADPTTYKPPRNKWDGIGHGEYVVIDFY
 GWLPDDLLRLCDRYPLTIVETKGTVPFLARSILITNSQTPLEWYSTAVPAVEALYR
 RITSLVFNKATQESTEGGQFVTLSPPCPEFPEYNY"
 674..679
 complement(704..1018)
 /note="predicted 11.9 kDa protein"
 /codon_start=1
 /product="ORF3"
 /protein_id="AAC35322.1"
 /db_xref="PID:g3598823"
 /db_xref="GI:3598823"
 /translation="MVTIPPLVSRMFPVCGFRVCKISPPFAFTTPRNPHNDVYISLPI
 TLLFFPAHFQFSQPAEISDKRYRVLNCHQTALAQGTHSSRQVTPLSLRSSSTL
 HQ"

polyA_signal
 CDS

CDS

complement(733..912)
 /note="predicted 6.5 kDa protein"
 /codon_start=1
 /product="ORF4"
 /protein_id="AAC35323.1"
 /db_xref="PID:g3598824"
 /db_xref="GI:3598824"
 /translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
 SNKVLTAVDRLSRCP"

CDS

complement(1035..1100)
 /note="predicted 2.3 kDa protein"
 /codon_start=1
 /product="ORF8"
 /protein_id="AAC35324.1"
 /db_xref="PID:g3598825"
 /db_xref="GI:3598825"
 /translation="MDIDHTVSDVDPRAASHKSHQ"

polyA_signal

CDS

complement(1336..1380)
 /note="predicted 1.8 kDa protein"
 /codon_start=1
 /product="ORF11"
 /protein_id="AAC35326.1"
 /db_xref="PID:g3598827"
 /db_xref="GI:3598827"
 /translation="MNNKNHYEVIKKTQ"
 complement(1369..1374)
 /note="predicted 27.8 kDa protein"
 /codon_start=1
 /product="ORF2"
 /protein_id="AAC35321.1"
 /db_xref="PID:g3598822"
 /db_xref="GI:3598822"

polyA_site

CDS

complement(1380..1767,1..314)
 /note="predicted 458 t
 BASE COUNT 447 a 360 c 502 g 458 t
 ORIGIN
 alignment_scores:
 Quality: 307.00 Length: 59
 Ratio: 5.293 Gaps: 0
 Percent Similarity: 98.305 Percent Identity: 96.610

alignment_block:

US-09-209-961-20 x AF055393

Align seg 1/1 to: AF055393 from: 1 to: 1767

1 MetYrThrSerLeuTrpGlyHisLeuGlyValValLysAlaasnGlyLe 17

```
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAAAATGGGCT 949
17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTGCAGACCCGGAACCCACATACCTGGAAACCCACTAGAAACAA 999
34 erGlyGlyMetValThrMetValLysLysTriLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAAGTGGTGTATTGATGACTTTT 1049
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCCTGGGATGATCTAC 1076
```

OM of: US-09-209-961-20 to: N_Geneseq_36.* out_format : pfs
Date: Dec 28, 1999 1:50 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=cdn2.1/uspt0.spool/us09209961/runat.22121999_101043_17942/app_query.fasta.1
-DB=N_Geneseq_36 -OFMT=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-CGAPOP=4.500 -CGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FCGAPOP=6.000 -FCGAPEXT=7.000 -FCGAPOP=10.000 -FCGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40.cdi -LIST=45 -DOCLALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-209-961-20
Query length: 59
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 140.030000

score_list:	Sequence	Strd Orig	zScore	EScore	Len	Documentation
	N_Geneseq_36:X35012	+	314.00	806.60	2.5e-37	1768 ! Genomic DNA sequence of PCV st
	N_Geneseq_36:X35213	+	314.00	806.60	2.5e-37	1768 ! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35212	+	314.00	806.60	2.5e-37	1768 ! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35214	+	314.00	806.60	2.5e-37	1768 ! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35210	+	307.00	787.91	2.8e-36	1767 ! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35211	+	307.00	787.91	2.8e-36	1767 ! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35013	+	114.00	272.33	1.4e-07	1759 ! Nucleotide sequence which has
	N_Geneseq_36:X41449	-	65.00	159.86	0.2716	315 ! Human secreted protein 5' Est S
	N_Geneseq_36:X06052	+	57.50	137.66	4.56	169 ! Human brain Expressed Sequence
	N_Geneseq_36:X40573	+	54.00	136.94	5.01	168 ! Human secreted protein 5' Est S
	N_Geneseq_36:X089836	+	55.50	134.48	6.86	309 ! Human death associated protein
	N_Geneseq_36:X46896	+	51.00	133.04	8.26	114 ! JESI-39D10 primer B2137CC. Huma
	N_Geneseq_36:T31291	-	63.50	130.83	10.96	3269 ! Rat poly-immunoglobulin recept
	N_Geneseq_36:X87541	-	54.50	130.74	11.09	342 ! Est clone DN293. New polynucleo
	N_Geneseq_36:X75294	-	58.00	128.92	14.01	980 ! Wheat ubiquitin-conjugating enz
	N_Geneseq_36:T25800	+	52.50	127.74	16.28	274 ! Human gene signature HUMGS08028
	N_Geneseq_36:T65664	+	50.00	124.95	23.30	190 ! Polymorphic repeat sequence cld
	N_Geneseq_36:T24240	+	53.00	123.77	27.11	452 ! Human gene signature HUMGS04028
	N_Geneseq_36:X78741	-	51.00	123.06	29.69	292 ! Staphylococcus aureus contig S
	N_Geneseq_36:X39589	-	51.50	123.00	29.91	333 ! Human secreted protein 5' Est S
	N_Geneseq_36:X86630	-	52.00	122.58	31.57	393 ! EST clone AS216. New polynucleo
	N_Geneseq_36:X39586	-	51.50	122.51	31.88	349 ! Human secreted protein 5' Est S
	N_Geneseq_36:X22390	-	47.50	121.07	38.00	145 ! Human HuN6 fragment for product
	N_Geneseq_36:T08171	+	56.50	121.04	38.34	1408 ! Full length coconut LPAAT clon
	N_Geneseq_36:T95297	+	58.00	120.84	40.49	2138 ! Rat glial cell line-derived ne
	N_Geneseq_36:X93311	+	58.00	120.54	40.49	2138 ! Glial cell-line derived neurot
	N_Geneseq_36:X08172	+	45.50	120.51	41.20	93 ! Primer for MMP19 coding sequence
	N_Geneseq_36:X39687	+	53.00	120.33	42.13	625 ! Gastric cancer associated gene
	N_Geneseq_36:X84975	+	58.00	119.51	46.79	2378 ! Rat glial cell derived neurot
	N_Geneseq_36:X054052	-	56.00	118.86	52.19	1557 ! Rat bone formation-inducing pr
	N_Geneseq_36:X062384	+	46.00	118.82	52.48	126 ! bioB gene regulatory region (p
	N_Geneseq_36:X003333	+	49.00	118.55	52.95	270 ! Fragment of post-transfusion, n
	N_Geneseq_36:X42651	-	49.00	118.55	52.95	270 ! PT-NANB virus lambda gtl1 clone
	N_Geneseq_36:X097712	-	51.00	118.40	53.96	453 ! Branched-chain alpha-ketoacid
	N_Geneseq_36:X20150	+	49.50	118.32	54.55	313 ! Probe (51) for microbial genes
	N_Geneseq_36:X90753	+	56.00	118.22	55.26	1624 ! Nucleotide sequence of clone X
	N_Geneseq_36:T73870	+	60.50	117.26	62.45	5518 ! Cotton fibre promoter clone 4-
	N_Geneseq_36:T73865	+	60.50	117.21	62.90	5547 ! Cotton fibre promoter clone 4-
	N_Geneseq_36:X41446	+	46.00	116.91	65.32	148 ! Human secreted protein 5' Est S
	N_Geneseq_36:T26682	-	48.50	116.64	67.65	285 ! Human gene signature HUMGS08932
	N_Geneseq_36:T18795	-	56.50	116.21	71.51	2226 ! Mouse immunophilin clone 213-1
	N_Geneseq_36:T69548	-	53.50	116.13	72.19	1053 ! Rat pheromone receptor VN5 cDN

N_Geneseq_36:N90105 - 53.00 116.12 72.35 930 ! Human non-specific cross-react
N_Geneseq_36:Q21401 + 47.50 115.75 75.85 241 ! Human leukocyte antigen DRw1
N_Geneseq_36:V43035 + 53.00 115.72 76.08 965 ! Streptococcus pneumoniae pol

seq_name: N_Geneseq_36:X35012

seq_documentation_block:
ID X35012 standard; DNA; 1768 BP.
AC X35012;
DT 01-JUL-1999 (first entry)
DE Genomic DNA sequence of PCV strain 999PCV.
KW PCV; strain PCV999; pig; PWS; porcine multisystemic wasting syndrome;
KW vaccine; ss.
OS Porcine circovirus
PN FR2769321-Al.
PD 09-APR-1999.
PF 03-OCT-1997; 012382.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PA (UYBE-) UNIV QUEENS BELFAST.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246947/21.
PT New porcine circovirus from animals with porcine systemic wasting
PT syndrome
PS Claim 13; Fig 1; 35pp; French.
CC The present sequence represents the genomic sequence of porcine
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
CC isolated from a physiological or tissue sample, particularly from
CC a lesion, from a pig showing symptoms of PWS (porcine multisystemic
CC wasting syndrome), or cultured cells, infected with PCV isolated from
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35012

Align seq 1/1 to: X35012 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValVallysAlaasnGlyLe 17
|||||
901 ATGTACACGTCATTTGGGGCCACCTGGTGGTGTGTAAGCAATGGGCT 950
|||||
17 uleuTleuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
951 GCTAAATTTTGCAGACCCGGAACCACTACTGGAAACCACTAGAACAA 1000
|||||
34 exGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1001 GTGGTGGGATGTTACCATGGTGAAGAGTGGTGTGTTATGATGACTTTT 1050
|||||
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1051 ATGGCTGGTGGTGGGATGATCTAC 1077

seq_name: N_Geneseq_36:X35213


```

seq_documentation_block:
ID X35213 standard; DNA; 1768 BP.
AC X35213.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl010.
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 4; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl010. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35213 ..
Align seg 1/1 to: X35213 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTCAGACCCGGAACACACATCTGGAAGCAACCACTAGAACAA 999

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36.X35212

seq_documentation_block:
ID X35212 standard; DNA; 1768 BP.
AC X35212.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 4; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl010. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35212 ..
Align seg 1/1 to: X35212 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTCAGACCCGGAACACACATCTGGAAGCAACCACTAGAACAA 999

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36.X35214

seq_documentation_block:
ID X35214 standard; DNA; 1768 BP.
AC X35214.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
of pregnant sows
PS Claim 14; Fig 3; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
Impl999. The specification describes a preparation of type II
porcine circovirus (PCV), which is particularly isolated from a lesion,
from a pig with symptoms of PMWS (porcine multisystemic wasting
syndrome). PCV (attenuated or inactivated), polypeptides derived from
it, and vectors that express these polypeptides are all useful in
vaccines, suitable for administration to adult or young pigs, or to
pregnant sows (for passive immunization of their offspring). DNA
isolated from PCV is used for in vivo or in vitro expression of viral
polypeptides, also as probes or primers for diagnosis in usual
hybridization or amplification assays. These polypeptides may also be
used diagnostically to detect PCV-specific antibodies, while antibodies
raised against the polypeptides can be used to detect antigens, in any
usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 314.00 Length: 59
Ratio: 5.322 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.305

alignment_block:
US-09-209-961-20 x X35212 ..
Align seg 1/1 to: X35212 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGTGTGTTAAAGCAAAATGGGCT 949

17 uLeuIleLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
|||||
950 GCTAATTTTCAGACCCGGAACACACATCTGGAAGCAACCACTAGAACAA 999

34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGGTTACCATGGTGAAGAGTGGTTGTTATTGATGACTTTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36.X35212

seq_documentation_block:
ID X35212 standard; DNA; 1768 BP.
AC X35212.
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN FR2769322-Al.

```

PS Claim 14; Fig 6; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Imp999. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:
 Quality: 314.00 Length: 59
 Ratio: 5.322 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.305
 alignment_block:
 US-09-209-961-20 x X35214 ..

Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
 901 ATGTACACGTCATTGCGGGCCACCTGGTGGTAAAGCAAAATGGGCT 950
 17 uLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
 951 GCTAATTTTGCAGACCCGGAACACCATCTGGAAACCACTAGAAACAA 1000
 34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
 1001 GTGGTGGGATGTTACCATGGTGAAGAAGTGGTTGTTATTGATGACTTTT 1050
 51 MetAlaGlyCysArgGlyMetIleTyr 59
 1051 ATGGCTGGCTGCGGTGGGATGATCTAC 1077
 seq_name: N_Geneseq_36:X35210

seq_documentation_block:
 ID X35210 standard; DNA; 1767 BP.
 AC X35210;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl011-48121.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 PI WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 of pregnant sows
 PS Claim 14; Fig 1; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl011-48121. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment_scores:
 Quality: 307.00 Length: 59
 Ratio: 5.293 Gaps: 0
 Percent Similarity: 98.305 Percent Identity: 96.610

alignment_block:
 US-09-209-961-20 x X35210 ..

Align seg 1/1 to: X35210 from: 1 to: 1767

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
 900 ATGTACACGTCATTGCGGGCCACCTGGTGGTAAAGCAAAATGGGCT 949
 17 uLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrS 34
 950 GCTAATTTTGCAGACCCGGAACACCATCTGGAAACCACTAGAAACAA 999
 34 erGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrPhe 50
 1000 GTGGTGGGATGTTACCATGGTGAAGAAGTGGTTGTTATTGATGACTTTT 1049
 51 MetAlaGlyCysArgGlyMetIleTyr 59
 1050 ATGGCTGGCTGCGGTGGGATGATCTAC 1076

seq_name: N_Geneseq_36:X35211

seq_documentation_block:
 ID X35211 standard; DNA; 1767 BP.
 AC X35211;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl011-48285.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 PI WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 of pregnant sows
 PS Claim 14; Fig 2; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl011-48285. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment_scores:
Quality: 307.00 Length: 59
Ratio: 5.293 Gaps: 0
Percent Similarity: 98.305 Percent Identity: 96.610

alignment_block:
US-09-209-961-20 x X35211 ..

Align seg 1/1 to: X35211 from: 1 to: 1767

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
|||||
900 ATGTACACGTCATTGTGGGCCACCTGGGTGTGTAAGCAAAATGGGCT 949
17 uLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThr 34
|||||
950 GCTAATTTTCAGACCCGGAACACACACTGGAACACCCATAGAACAA 999
34 eRgLYGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
1000 GTGGTGGGATGTTACCATCGTGAAGAAGTGTGTTATTGATGACITTT 1049

51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
1050 ATGGCTGGCTGCCCTGGGATGATCTAC 1076

seq_name: N_Geneseq_36.X35013

seq_documentation_block:
ID X35013 standard; DNA; 1759 BP.
AC X35013;
DE 01-JUL-1999 (first entry)
KW Nucleotide sequence which has homology to PCV sequence.
KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
KW vaccine; ss.
OS Sus sp.
PN FR2769321-AL.
PD 09-APR-1999.
PF 03-OCT-1997; 012382.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PA (UYBE-) UNIV QUEENS BELFAST.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246947/21.
PT New porcine circovirus from animals with porcine systemic wasting syndrome
PS Disclosure: Fig 2; 35pp; French.
CC The specification describes a genomic sequence of porcine
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
CC isolated from a physiological or tissue sample, particularly from
CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
CC wasting syndrome), or cultured cells, infected with PCV isolated from
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format. The present sequence appears in the
CC specification.
SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment_scores:
Quality: 114.00 Length: 59
Ratio: 2.923 Gaps: 0
Percent Similarity: 66.102 Percent Identity: 47.458

alignment_block:
US-09-209-961-20 x X35013/rev ..

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

1 MetTyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLe 17
:::|||||
872 CTGTACACGTCATAGTGGCCGCCCGGTTGTGGGAAGAGCCAGTGGGCC 823
17 uLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThr 34
:::|||||
822 CGTAATTTTGTGCTAGCCCTAGGACACCTACTGGAAGCCCTAGTAAATAA 773
34 eRgLYGlyMetValThrMetValLysLysTrpLeuLeuMetThrPhe 50
|||||
772 GTGGTGGGATGATCATCGGGAAGAAGTGTGTTTGGATGATTTT 723
51 MetAlaGlyCysArgGlyMetIleTyr 59
|||||
722 ATGGCTGGTTACCTGGGATGATCTAC 696

seq_name: N_Geneseq_36.X41449

seq_documentation_block:
ID X41449 standard; cDNA; 315 BP.
AC X41449;
DE 22-JUN-1999 (first entry)
KW Human secreted protein 5; EST SEQ ID NO: 108 from WO 9906553.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; antiinflammatory; tumour inhibition; antitumour; ds.
OS Homo sapiens.
PN WO9906553-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1237.
PR 01-AUG-1997; US-905051.
PA (GEST) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153783/13.
DR P-PSDB; Y12591.
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from umbilical cord, lymph ganglia,
PT lymphocytes and placental tissue
PS Claim 1; Page 219-220; 41pp; English.
CC X41379 to X41526 represent 5, expressed sequence tags (ESTs) for human
CC secreted proteins, and encode the proteins given in Y12521 to Y12668,
CC respectively. The proteins given represent the signal peptide and an
CC N-terminal fragment of a secreted protein. The nucleic acid sequences
CC can be used for producing secreted human gene products. They can also
CC be used to develop products for diagnosis and therapy. The proteins
CC obtained may have cytokine activity, cell proliferation/differentiation
CC activity, haematopoiesis regulating activity, tissue growth regulating
CC activity, reproductive hormone regulating activity, chemotactic/
CC chemokinetic activity, haemostatic and thrombolytic activity, receptor/
CC ligand activity, antiinflammatory activity, tumour inhibition activity
CC or other activities. The products can be used in forensic, gene therapy
CC and chromosome mapping procedures. The sequences can also be used for
CC obtaining corresponding promoter sequences. The nucleic acids encoding
CC the signal peptide can be used for directing extracellular secretion of
CC a polypeptide or the insertion of a polypeptide into a membrane, or
CC importing a polypeptide into a cell.
SQ Sequence 315 BP; 67 A; 85 G; 71 T;

alignment_scores:
Quality: 65.00 Length: 42
Ratio: 2.500 Gaps: 1
Percent Similarity: 61.905 Percent Identity: 35.714

alignment_block:

```

US-09-209-961-20 x X41449/rev ..
Align seg 1/1 to reverse of: X41449 from: 1 to: 315

      7 GlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnThrAr 23
      289 GGCCACTGTGGATCTGTCCTCTGGAGGCCATCTGTCTGCCGCCACGGG 240
      23 glysProHisThrGlyAsnHis.....LeuLysThrSerGlyGlyMetV 38
      1:||||| :|||||:||||| :|||||:|||||:|||||:|||||:|||||
      239 CCGTCCATGGTCAGGACGACACCCACTGASTGACGCTGAGGGGGCTACAG 190
      38 alThrMetValLysLysTrpLeuLeu 46
      1:||||| :|||:|||||
      189 TGAACCTGGACAGCTTGGCTCTCG 164

seq_name: N_Geneseq_36:Q60652

seq_documentation_block:
ID Q60652 standard; cDNA; 379 BP.
AC Q60652;
DT 16-MAR-1994 (first entry)
DE Human brain Expressed Sequence Tag EST02666.
KW Gene transcription product; genetic markers; tagging; in vivo;
OS Homo sapiens.
PN WO9316178-A.
PD 19-AUG-1993.
PF 12-FEB-1993; U01294.
PR 12-FEB-1992; US-837195.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICE.
PI Adams MD, Moreno RF, Venter CJ;
DR WPI: 93-272882/34.
PT Enriched oligonucleotides and corresp. sequences - used as
PT markers for human genes transcribed in-vivo, facilitate tagging
PT of most human genes
PS Example 4; Page 368; 500pp; English.
CC The Expressed Sequence Tag was isolated from a human brain cDNA
CC library as part of a large set of ESTs which can be used as markers
CC for human genes transcribed in vivo. They can be used to facilitate
CC tagging of most human genes, for mapping locations of expressed genes
CC on chromosomes, for individual or forensic identification, for mapping
CC locations of disease-associated genes, for identification of tissue
CC type, and for prep. of antisense sequences, probes and constructs.
CC EST02666 has a "poor" coding probability as evaluated using the
CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 379 BP; 70 A; 117 C; 76 G; 110 T;

alignment_scores:
      Quality: 57.50      Length: 36
      Ratio: 2.614      Gaps: 1
      Percent Similarity: 61.111      Percent Identity: 38.889

alignment_block:
US-09-209-961-20 x Q60652 ..
Align seg 1/1 to: Q60652 from: 1 to: 379

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
38 CTCCTTCGCTCCAGCCACATCACCAGCCGCCGAGTCCACCTAGCACC 87
33 rSerGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThrP 50
||| ||| :|||:|||||
88 TTCCTCGGATNATCTCC.....CCCTGGTGGCTCTTCTACTT 128

50 heMetAla 52
: |||
129 ATTCAGCC 136

seq_name: N_Geneseq_36:X40573

```

```

seq_documentation_block:
ID X40573 standard; cDNA; 168 BP.
AC X40573;
DT 18-JUN-1999 (first entry)
DE Human secreted protein 5', EST SEQ ID No: 173.
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition; ds.
QS Homo sapiens.
PN WO9906550-A2.
PD 11-FEB-1999.
PF 31-JUL-1998; IB1232.
PR 01-AUG-1997; US-905144.
PA (GEST ) GENSET.
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
DR WPI; 99-153780/13.
DR P-PSDB: Y11851.
PT New isolated prostate-derived nucleic acids - used to develop
PT products which may have cytokine, immune regulatory, haematopoiesis
PT regulating, anti-inflammatory or tumour inhibition activity
PS Claim 1; Page 276; 675pp; English.
CC X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human
CC secreted proteins expressed in prostate, and encode the proteins given in
CC peptide and an N-terminal fragment of a secreted protein. The nucleic
CC acid sequences can be used for producing secreted human gene products.
CC They can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell proliferation and
CC differentiation activity, haematopoiesis regulating activity, tissue
CC growth regulating activity, reproductive hormone regulating activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
CC receptor/ligand activity, anti-inflammatory activity, tumour inhibition
CC activity or other activities. The products can be used in forensic, gene
CC therapy and chromosome mapping procedures. The sequences can also be used
CC for obtaining corresponding promoter sequences. The nucleic acids
CC encoding the signal peptides can be used for directing extracellular
CC secretion of a polypeptide or the insertion of a polypeptide into a
CC membrane, or importing a polypeptide into a cell.
SQ Sequence 168 BP; 61 A; 27 C; 39 G; 39 T;

alignment_scores:
      Quality: 54.00      Length: 33
      Ratio: 2.455      Gaps: 0
      Percent Similarity: 66.667      Percent Identity: 33.333

alignment_block:
US-09-209-961-20 x X40573/rev ..
Align seg 1/1 to reverse of: X40573 from: 1 to: 168

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuGlnTh 22
||| ||||| :|||:|||||:|||||:|||||:|||||:|||||
102 TGGCAGCATCTTCAATACTGTAATCTGTCTGTGGCTCTCTGAAGACGATGAT 53

22 rArgLysProHisThrGlyAsnHisThrSerGlyGlyMetVal 38
||||| :|||:|||||:|||||:|||||:|||||:|||||:|||||
52 GAGGAGAGAGGAGAACATAGGATCATCTGAAGAAAAGACGTCACCTACTGATC 4

seq_name: N_Geneseq_36:Q89836

seq_documentation_block:
ID Q89836 standard; cDNA; 309 BP.
AC Q89836;
DT 04-DEC-1995 (first entry)
DE Human death associated protein DAP-1.
KW Death associated protein; DAP; cytokine; cell death; ss.
OS Homo sapiens.
PF Key Location/Qualifiers
FT cds 1..309

```

```

FT W09510630-A. /*tag= a
PN 20-APR-1995.
PD 12-OCT-1994; U11598.
PF 12-OCT-1993; IL-107250.
PR (RYCUV) RYCUS A.
PA (YEDA ) YEDA RES & DEV CO LTD.
PI Kimchi A;
DR WPI; 95-178528/23.
DR P-PSDB; R74204.
DR DNA whose expression mediates cytokine-induced programmed cell
PT death - used to treat diseases or disorders associated with
PT uncontrolled, pathological cell growth or cytokine-induced
PT programmed cell death.
PS Claim 2; Fig 6; 61pp; English.
CC DAP genes seem to play an imp. role in programmed cell death and the
CC inhibition of their expression protects the cell from cytokine-
CC promoted cell death. An HL-60 cDNA library constructed in lambda
CC gt10 vector was screened with the cDNA insert of pTK01-230. Two
CC independent clones, lambda2 and lambda2a, almost completely
CC overlapping and carrying cDNA inserts of about 2.3 kb were analysed.
CC Lambda2 cDNA clone encompasses the 5'-UTR, short coding region(s)
CC and a relatively long 3' UTR that constitutes more than 60% of the
CC cDNA clone. The nt. sequence of the cDNA carried by lambda2 and
CC predicted AA pattern are present in Q89835/R74204. The ORF is
CC preceded by an extremely GC-rich 5' UTR and potentially codes for
CC a protein consisting of 102 AAs with calculated MW of 11.2 kDa.
CC It has an isoelectric pt. of 10 and is rich in prolines. Q89836
CC is the DAP-1 ORF.
SQ Sequence 309 BP; 90 A; 96 C; 79 G; 44 T;

alignment_scores:
Quality: 55.50 Length: 23
Ratio: 3.083 Gaps: 1
Percent Similarity: 78.261 Percent Identity: 52.174

alignment_block:
US-09-209-961-20 x Q89836 ..
Align seg 1/1 to: Q89836 from: 1 to: 309

7 GlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuLeuGlnThrAr 23
||||| :|||:|||||:|||||:|||||:|||||:
40 GGACACCCGCCCGCTGAAGCTGGTGAATCGAATTCGCAG...AA 86

23 GlyProHisThrGlyAsn 29
: |||||:|||||:
87 ACACCCACATACAGGAGAC 105

seq_name: N_Geneseq_36:Q46896

seq_documentation_block:
ID Q46896 standard; DNA; 114 BP.
AC Q46896;
DT 14-FEB-1994 (first entry)
DE JES1-39D10 primer B2137CC.
KW Heavy; light; chain; anti-human; monoclonal; antibody; JES1-39D10;
KW complementarity determining region; CDR; loop analysis; IL-5; human;
KW Kabat determination; grafting; PCR; polymerase chain reaction; amplify;
KW primer; ss.
OS Synthetic.
PN W09516184-A.
PD 19-AUG-1993.
PF 04-FEB-1993; U00759.
PR 06-FEB-1992; US-832842.
PA (SCHE ) SCHERING CORP.
PI Abrams JS, Chou C, Jenh C, Murgolo NJ, Petro ME;
PI Silver JE, Tindall S, Windsor WT, Zavodny PJ;
DR WPI; 93-272888/34.
DR Humanised monoclonal antibody - comprises variable animal region
PT and constant human region, binds to human interleukin-5
PS Example; Page 71; 118pp; English.

```

```

CC The sequences given in Q46892-99 are primers which were used in the
CC amplification of the heavy and light chains of the anti-human mono-
CC clonal antibody JES1-39D10. These primer sequences were based on
CC the IgG2a/kappa isotype of JES1-39D10. The amplification products
CC were isolated by gel electrophoresis and cloned separately into pUC19.
CC The ligated plasmids were used to transform E. coli DH5-alpha. The
CC position of the complementarity determining regions (CDRs) in the
CC amplified and expressed sequences, was determined by the method of
CC Kabat et al and by loop analysis. These CDRs may be used to produce
CC humanised antibodies by grafting them onto a human antibody. Amino
CC acid residues from outside the CDR may also be included in the graft
CC if they interact with the CDRs or with IL-5.
SQ Sequence 114 BP; 27 A; 21 C; 33 G; 33 T;

alignment_scores:
Quality: 51.00 Length: 20
Ratio: 3.643 Gaps: 0
Percent Similarity: 70.000 Percent Identity: 55.000

alignment_block:
US-09-209-961-20 x Q46896/rev ..
Align seg 1/1 to reverse of: Q46896 from: 1 to: 114

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| :|||:|||||:|||||:|||||:
105 CTCTCCAGATTCACACAGCTTCACCTCACACTGGATACCAATTTAAAGT 56

33 rSerGlyGly 36
: |||
55 GTTAAAGGA 46

seq_name: N_Geneseq_36:T31291

seq_documentation_block:
ID T31291 standard; cDNA; 3269 BP.
AC T31291;
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor, cDNA.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's I3 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus; ss.
OS Rattus rattus.
FH Key
FT cds
FT Location/Qualifiers
/*tag= a
PN W09621012-Al.
PD 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI; 96-333987/33.
DR P-PSDB; W03181.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 123-127; 152pp; English.
CC The present sequence encodes the rat poly-immunoglobulin (Ig)
CC receptor, a portion of which corres. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal

```



```

seq_name: /cgn2_6/ptodata/1/1na/5D_COMB.seq:US-08-327-451E-21
seq_documentation_block:
; Sequence 21, Application US/08327451E
; Patent No. 5910630
; GENERAL INFORMATION:
; APPLICANT: Davies, Maelor
; APPLICANT: Hawkins, Deborah
; APPLICANT: Nelson, Janet
; APPLICANT: Lassen, Michael
; TITLE OF INVENTION: PLANT LYSOPHOSPHATIDIC
; TITLE OF INVENTION: ACID ACYLTRANSFERASES
; NUMBER OF SEQUENCES: 37
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB
; COMPUTER: IBM PC
; OPERATING SYSTEM: Windows NT 4.0
; SOFTWARE: Microsoft Word For Windows 7.0a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/327,451E
; FILING DATE: 21-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/254,404
; FILING DATE: 06-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/231,196
; FILING DATE: 21-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/224,625
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 106-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (530) 753-6313
; TELEFAX: (530) 753-1510
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-327-451E-21

alignment_scores:
Quality: 56.50 Length: 66
Ratio: 1.614 Gaps: 2
Percent Similarity: 53.030 Percent Identity: 22.727

alignment_block:
US-09-209-961-20 x US-08-327-451E-21
Align seg 1/1 to: US-08-327-451E-21 from: 1 to: 1408
2 TyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLe 18
||| ||||| ||| |||:||||| |||
248 TATATAGCTCATGTGCTTCAGGGCAAGTTCGTTCTTCGGGGCCGT 297
18 u1leuGln.....ThrArgLysp 25
|||: |||
298 TGCTGGAGAGCTCTCAAGCGAGCTTCGGGATGTCACCAACCGAAGA 347
25 roHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValThrMetVal 41
||| |||||:||||| |||
348 TGCAGCGGGCAACCGAGTC.....GCCGGCCGCCGCGCGGATG 388
42 LysLysTrpLeuLeuMetThrPheMetAlaGlyCysArgGlyMet 57
|||:|||||:|||||:|||||:|||||
389 ACTTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 436
; TITLE OF INVENTION: PLANT LYSOPHOSPHATIDIC
; TITLE OF INVENTION: ACID ACYLTRANSFERASES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Calgene, Inc.
; STREET: 1920 Fifth Street
; CITY: Davis
; STATE: CA
; COUNTRY: USA
; ZIP: 95616
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.0 MB
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/254,404
; FILING DATE: 06-JUN-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/231,196
; FILING DATE: 21-APR-1994
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/224,625
; FILING DATE: 06-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Elizabeth Lassen
; REGISTRATION NUMBER: 31,845
; NAME: Donna E. Scherer
; REGISTRATION NUMBER: 34,719
; NAME: Carl J. Schwedler
; REGISTRATION NUMBER: 36,924
; REFERENCE/DOCKET NUMBER: CGNE 106-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (916) 753-6313
; TELEFAX: (916) 753-1510
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1408 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; US-08-254-404-21

alignment_scores:
Quality: 56.50 Length: 66
Ratio: 1.614 Gaps: 2
Percent Similarity: 53.030 Percent Identity: 22.727

alignment_block:
US-09-209-961-20 x US-08-254-404-21
Align seg 1/1 to: US-08-254-404-21 from: 1 to: 1408
2 TyrThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLe 18
||| ||||| ||| |||:||||| |||
248 TATATAGCTCATGTGCTTCAGGGCAAGTTCGTTCTTCGGGGCCGT 297
18 u1leuGln.....ThrArgLysp 25
|||: |||
298 TGCTGGAGAGCTCTCAAGCGAGCTTCGGGATGTCACCAACCGAAGA 347
25 roHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValThrMetVal 41
||| |||||:||||| |||
348 TGCAGCGGGCAACCGAGTC.....GCCGGCCGCCGCGCGGATG 388
42 LysLysTrpLeuLeuMetThrPheMetAlaGlyCysArgGlyMet 57
|||:|||||:|||||:|||||:|||||
389 ACTTGTGGATGATGATGATGATGATGATGATGATGATGATGATGATG 436

```

[illegible]

... |||||...: ||| :
348 TGCAGCGGGCAACCGAGTC.....GCCGGCGGCGGACGGGATG 388

42 LysLysTrpLeuLeuMetThrPheMetAlaGlyCysArgGlyMet 57
|||||...: |||||...:

389 ACTTGTGGATGATGATAGATGATTACTGCTGCTGCGTGGTTA 436

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-148-910-10

seq_documentation_block:

; Sequence 10, Application US/08148910
; Patent No. 5466593

; GENERAL INFORMATION:

; APPLICANT: Takeshi SHIMOMURA et al.

; TITLE OF INVENTION: NO. 5466593el Protein and Gene Encoding Said Protein

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch,

; MEDIUM TYPE: 500 Kb Storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/148,910

; FILING DATE: No. 5466593ember 5, 1993

; CLASSIFICATION: 435

; PRIOR APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-8850

; TELEFAX: 202-371-8856

; TELEX:

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 71 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Other nucleic acid (synthetic DNA)

; ORIGINAL SOURCE:

; ORGANISM: human

; US-08-148-910-10

alignment_scores:

Quality: 44.00 Length: 13

Ratio: 3.667 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 53.846

alignment_block:

US-09-209-961-20 x US-08-148-910-10 ..

Align seg 1/1 to: US-08-148-910-10 from: 1 to: 71

46 LeuLeuMetThrPheMetAlaGlyCysArgGlyMetIle 58

|||||...: |||||...: |||

33 TTGCTGTCACATCTCTGGCAGGATGCGAGGCCAAGGTG 71

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-448-937A-10

seq_documentation_block:

; Sequence 10, Application US/08448937A

; Patent No. 5677164

; GENERAL INFORMATION:

; APPLICANT: Takeshi SHIMOMURA et al.

; TITLE OF INVENTION: NO. 5677164el Protein and Gene Encoding Said Protein

; NUMBER OF SEQUENCES: 14

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Wenderoth, Lind & Ponack

; STREET: 805 Fifteenth Street, N.W., #700

; CITY: Washington

; STATE: D.C.

; COUNTRY: U.S.A.

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 5.25 inch,

; MEDIUM TYPE: 500 Kb Storage

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Wordperfect

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/448,937A

; FILING DATE: May 24, 1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/148,910

; FILING DATE: No. 5677164ember 5, 1993

; ATTORNEY/AGENT INFORMATION:

; NAME: Warren M. Cheek, Jr.

; REGISTRATION NUMBER: 33,367

; REFERENCE/DOCKET NUMBER:

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-8850

; TELEFAX: 202-371-8856

; TELEX:

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 71 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: Other nucleic acid (synthetic DNA)

; ORIGINAL SOURCE:

; ORGANISM: human

; US-08-448-937A-10

alignment_scores:

Quality: 44.00 Length: 13

Ratio: 3.667 Gaps: 0

Percent Similarity: 92.308 Percent Identity: 53.846

alignment_block:

US-09-209-961-20 x US-08-448-937A-10 ..

Align seg 1/1 to: US-08-448-937A-10 from: 1 to: 71

46 LeuLeuMetThrPheMetAlaGlyCysArgGlyMetIle 58

|||||...: |||||...: |||

33 TTGCTGTCACATCTCTGGCAGGATGCGAGGCCAAGGTG 71

seq_name: /cgn2_6/ptodata/1/lna/backfiles1.seq:5218099-4

seq_documentation_block:

; Patent No. 5218099

; APPLICANT: REYES, GREGORY R.;BRADLEY, DANIEL W.;RABIN,LINDA;

; FRY, KIRK

; TITLE OF INVENTION: POST-TRANSFUSION, NON-A, NON-B HEPATITIS

; VIRUS POLYNUCLEOTIDES

; NUMBER OF SEQUENCES: 17

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/372,711

; FILING DATE: 28-JUN-1989

; PRIOR APPLICATION DATA:

```
; APPLICATION NUMBER: 350,570; 334,701; 228,334; 215,728; 846,757
; FILING DATE: 11-MAY-1989
; APPLICATION NUMBER: 334,701
; FILING DATE: 06-APR-1989
; APPLICATION NUMBER: 228,334
; FILING DATE: 04-AUG-1988
; APPLICATION NUMBER: 215,728
; FILING DATE: 06-JUL-1988
; APPLICATION NUMBER: 846,757
; FILING DATE: 01-APR-1986
; SEQ ID NO: 4
; LENGTH: 270
5218099-4
```

```
alignment_scores:
  Quality: 49.00      Length: 37
  Ratio: 1.885       Gaps: 1
Percent Similarity: 70.270 Percent Identity: 29.730

alignment_block:
US-09-209-961-20 x 5218099-4/rev ..
Align seg 1/1 to reverse of: 5218099-4 from: 1 to: 270

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| : : : : : ||||| : : : : : ||||| : : : : :
130 TTACTTTATGTAGCGAATGCGCATCTTGCTCACATAGCGGTC 81
: : : : : : : : : : : : : : : : : : : : :
33 rSerGlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetThr 50
: : : : : : : : : : : : : : : : : : : : :
80 GAAG.....CAATCGAGCTTGCGCGCATATTGCCACTTTTGGT 37

50 heMetAlaGly 53
||||| : : : : :
36 TTCAGTTGGC 26
```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-470-892-4

```
seq_documentation_block:
; Sequence 4, Application US/08470892
; Patent No. 5707839
; GENERAL INFORMATION:
; APPLICANT: DENOYA, CLAUDIO D.
; APPLICANT: STUTZMAN-ENGWALL, KIM J.
; TITLE OF INVENTION: GENES ENCODING BRANCHED-CHAIN
; TITLE OF INVENTION: ALPHA-KETOACID DEHYDROGENASE COMPLEX FROM STREPTOMYCES
; TITLE OF INVENTION: AVERMITILIS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DR. PETER C. RICHARDSON, PFIZER INC
; STREET: 235 EAST 42ND STREET, 20TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10017-5755
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,892
; FILING DATE: 6-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: SHEYKA, ROBERT F.
; REGISTRATION NUMBER: 31,304
; REFERENCE/DOCKET NUMBER: PC8529B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)573-1189
; TELEFAX: (212)573-1939
; TELEX: N/A
```

```
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 453 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-470-892-4

alignment_scores:
  Quality: 51.00      Length: 42
  Ratio: 1.700       Gaps: 3
Percent Similarity: 71.429 Percent Identity: 33.333

alignment_block:
US-09-209-961-20 x US-08-470-892-4/rev ..
Align seg 1/1 to reverse of: US-08-470-892-4 from: 1 to: 453

17 LeuLeuLeuLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysTh 33
||||| : : : : : ||||| : : : : : ||||| : : : : :
432 CTTCGGCTGCTCCAGGATCGCGC...CACGTCGGCCACGACCTT.... 391
: : : : : : : : : : : : : : : : : : : : :
33 rSerGlyGlyMetValThrMet.ValLysLysTrpLeuLeuLeuMetThr 49
||||| : : : : : ||||| : : : : : ||||| : : : : :
390 ....GGAGCCAGCTCGCGTCGACGAGCGGATGTCGAAGCTGACCGCC 345

50 heMetAlaGlyCysArgGlyMet 57
: : : : : : : : : : : : : : : : : : : : :
344 AGCGTGGTGACCTGTCGGGGCTTG 321

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-2

seq_documentation_block:
; Sequence 2, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
```

```
;
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-2

alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732      Gaps: 4
  Percent Similarity: 50.000  Percent Identity: 33.929

alignment_block:
US-09-209-961-20 x US-08-050-073-2  ..
Align seg 1/1 to: US-08-050-073-2 from: 1 to: 269

3 ThrSerLeuTrrpGlyHisLeuGlyValVallysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGCCTGAT 158
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
||| |||||:|:|:|
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193

35 .....GlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetTh 49
||||| ||||| :||| |||||
194 GCAGAGCGGG.....CCGGGTGGACACCTATTGCGAGAC 228

49 rPheMetAlaGlyCys 54
| :|:| |||||
229 ACAACTAGCGGGCTGT 244
```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-050-073-9

```
seq_documentation_block:
; Sequence 9, Application US/080500073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-9

alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732      Gaps: 4
  Percent Similarity: 50.000  Percent Identity: 33.929

alignment_block:
US-09-209-961-20 x US-08-050-073-9  ..
Align seg 1/1 to: US-08-050-073-9 from: 1 to: 269

3 ThrSerLeuTrrpGlyHisLeuGlyValVallysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGCCTGAT 158
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
||| |||||:|:|:|
159 G.....CCGAGTACTGGAACAGCCAGAGGACCTCTCTGGA 193

35 .....GlyGlyMetValThrMetValLysLysTrpLeuLeuLeuMetTh 49
||||| ||||| :||| |||||
194 GCAGAGCGGG.....CCGAGGTGGACACCTACTGCGAGAC 228

49 rPheMetAlaGlyCys 54
| :|:| |||||
229 ACAACTAGCGGGTGT 244

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-050-073-10

seq_documentation_block:
; Sequence 10, Application US/080500073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRBeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 10:
; INFORMATION CHARACTERISTICS:
```

```
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-10
```

```
alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732        Gaps: 4
Percent Similarity: 50.000 Percent Identity: 33.929
```

```
alignment_block:
```

```
US-09-209-961-20 x US-08-050-073-10 ..
```

```
Align seg 1/1 to: US-08-050-073-10 from: 1 to: 269
```

```
3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCTGAT 158
: ||| ||||| : ||| |||||
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
: ||| ||||| : ||| |||||
159 G.....CGAGTACTGGAAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetTh 49
||||| ||||| ||||| ||||| |||||
194 GCAGAGCGGGG.....CGCGGTGGACACCTACTGTCAGAC 228
49 rPheMetAlaGlyCys 54
| :||| |||||
229 ACAACTACGGGGTTGT 244
```

```
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-050-073-12
```

```
seq_documentation_block:
```

```
; Sequence 12, Application US/08050073
; Patent No. 5567809
; GENERAL INFORMATION:
; APPLICANT: Apple, Raymond J.
; APPLICANT: Begovich, Ann B.
; APPLICANT: Bugawan, Teodorica L.
; APPLICANT: Erlich, Henry A.
; APPLICANT: Griffith, Robert L.
; APPLICANT: Scharf, Stephen J.
; TITLE OF INVENTION: Methods and Reagents for HLA DRbeta DNA
; TITLE OF INVENTION: Typing
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2974
; TELEFAX: (510) 814-2977
```

```
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-050-073-12
```

```
alignment_scores:
  Quality: 48.50      Length: 56
  Ratio: 1.732        Gaps: 4
Percent Similarity: 50.000 Percent Identity: 33.929
```

```
alignment_block:
```

```
US-09-209-961-20 x US-08-050-073-12 ..
```

```
Align seg 1/1 to: US-08-050-073-12 from: 1 to: 269
```

```
3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu11 19
||||| ||||| ||| :||| |||
109 ACAGCGACGTGGGGAGTACCGGGCGGTGACGGAGCTGGGGCGGCTGAT 158
: ||| ||||| : ||| |||||
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSer.... 34
: ||| ||||| : ||| |||||
159 G.....CGAGTACTGGAAACAGCCAGAGGACCTCTCTGGA 193
35 .....GlyGlyMetValThrMetValLysLysTrpLeu.LeuLeuMetTh 49
||||| ||||| ||||| ||||| |||||
194 GCAGAGCGGGG.....CGCGGTGGACACCTACTGTCAGAC 228
49 rPheMetAlaGlyCys 54
| :||| |||||
229 ACAACTACGGGGTTGT 244
```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-264-250A-2
```

```
seq_documentation_block:
```

```
; Sequence 2, Application US/08264250A
; Patent No. 5663047
; GENERAL INFORMATION:
; APPLICANT: Fumiyu OBATA, et al.
; TITLE OF INVENTION: NOVEL HLA-DR ANTIGEN GENE AND ITS NUCLEOTIDE
; TITLE OF INVENTION: SEQUENCE AND ITS USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Law Office of Sherman and Shalloway
; STREET: 413 N. Washington Street
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22314
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
; COMPUTER: IBM Clone, 8088 Turbo
; OPERATING SYSTEM: MS DOS 5.0
; SOFTWARE: Word Perfect, Version 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/264,250A
; FILING DATE: 22-JUN-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/745,444
; FILING DATE: August 15, 1991
; APPLICATION NUMBER: JP A00217209/1990
; FILING DATE: August 20, 1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Richard A. Steinberg
; REGISTRATION NUMBER: 26,588
; REFERENCE/DOCKET NUMBER: HIR-109
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 549-2282
```



```

;
; NUMBER OF SEQUENCES: 315
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hoffmann-La Roche Inc.
; STREET: 340 Kingsland Street
; CITY: Nutley
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/050,073
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Petry, Douglas A.
; REGISTRATION NUMBER: 35,321
; REFERENCE/DOCKET NUMBER: 8769
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 814-2977
; TELEFAX: (510) 814-2977
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 269 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-050-073-8

```

```

alignment_scores:
    Quality: 47.50          Length: 53
    Ratio: 1.759           Gaps: 3
    Percent Similarity: 50.943    Percent Identity: 33.962

alignment_block:
US-09-209-961-20 x US-08-050-073-8 ..

Align seg 1/1 to: US-08-050-073-8 from: 1 to: 269

3 ThrSerLeuTrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeu1 19
||||| ||||| ||| :||| |||:
109 ACAGCGACGTGGGGAGTACCGGGGTGACGGGCTGGGGCGGCTGAT 158
19 eLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerGlyG 36
: ||| |||||: :|||
159 G.....CCGAGTACTGGAACAGCCAGAGACATCC....189

36 1yMetValThrMetValLysLysTrpLeu.LeuLeuMetThrPheMetA1 52
||| ||| |||||: ||| :|||
190 ..TGAAGACGAGCGCGCGGTGGACACCTACTGCAGACACACTACG 237

52 aGlyCys 54
|||||
238 GGGTTGT 244

```


2


```

source
1..152
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="1341694"
/clone_lib="Soares.parathyroid_tumor_NbHPA"
/tissue_type="parathyroid tumor"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: parathyroid gland; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer"
[5-
TGTACCAATCTGAAGTGGAGCGCGCCACCAATTTTTTTTTTTTTTTTTTTTTT
T-3'], double-stranded cDNA was size selected, ligated to
Eco RI adapters (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of a modified pT7T3
vector (Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
adenomas was kindly provided by Dr. Stephen Marx, National
Institute of Diabetes and Digestive and Kidney Diseases,
NIH."
BASE COUNT      37 a  42 c  36 g  37 t
ORIGIN

alignment_scores:
    Quality:      59.00      Length:      34
    Ratio:        2.458      Gaps:      1
    Percent Similarity: 70.588      Percent Identity: 38.235

alignment_block:
US-09-209-961-20 x AA723992/rev ..
Align seg 1/1 to reverse of: AA723992 from: 1 to: 152

6 TrpGlyHisLeuGlyValVallyAlaAsnGlyLeuLeu..... 18
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 TGGCACCATCTGGGCTGTGTCAGGCGCCATGGAGGTGGAAGGCCACCA 99

19 .lLeuGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerG 35
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 CATCCTTAAGCCATCAGTAGTATAGTGGTGCCACCCTGCATGTGAAGG 49

35 ly 35
||
48 GG 47

seg_name: gb_gss3:AQ090205

seg documentation_block:
LOCUS AQ090205 430 bp DNA GSS 25-AUG-1998
DEFINITION HS_3009_AL_F06_T7 CIT Approved Human Genomic Sperm Library D Homo
sapiens genomic clone Plate-3009 Col-11 Row-K, genomic survey
sequence.
ACCESSION AQ090205
NID G3459116
VERSION AQ090205.1 GI:3459116
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 430)
Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
Construction of a Characterized Clone Resource for Genomic
Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
Tagged Connectors
unpublished (1997)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L

```

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40UP from Gibco.

FEATURES

source
1. .278
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1854731"
/clone_lib="Soares_NFL_T_GBC_S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not 1; Site_2: Eco RI; Equal amounts of plasmid DNA from three normalized libraries (fetal lung NBHL19W, testis NHT, and B-cell NCI-GAP GCBI) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 82 a 41 c 56 g 99 t
ORIGIN
alignment_scores:
Quality: 60.50 Length: 40
Ratio: 2.420 Gaps: 1
Percent Similarity: 62.500 Percent Identity: 32.500

alignment_block:

US-09-209-961-20 x AI283571/rev ..

Align seg 1/1 to reverse of: AI283571 from: 1 to: 278

```

4 SerLeuTrpGlyHisLeuGlyValVallylsAlaAsnGlyLeuLeuLe 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
117 TCCTATGAGCCACACCGACAGTACATGCT.....AT 77

20 uGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGly 37
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
76 TTATTCAGGATCCCGACGAGAGCCATGTAACCTATAAAGCTA 27

37 etValThrMetVallyls 43
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
26 TAATTCATTCTTAAAAAAA 7

```

seq_name: gb_est34:AV139690

seq_documentation_block: 284 bp mRNA EST 02-JUL-1999
LOCUS AV139690 Mus musculus C57BL/6J 10-11 day embryo Mus musculus cDNA
DEFINITION clone 2810052C23, mRNA sequence.

ACCESSION AV139690

NID 95343685

VERSION AV139690.1 GI:5343685

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

1 (bases 1 to 284) Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

AUTHORS Carninci, P., Shibata, K., Ozawa, Y., Konno, H., Itoh, M., Aizawa, K.,

Akaiira, S., Akiyama, J., Fukuda, S., Fukunishi, Y., Funayama, T.,

Hara, A., Hayatsu, N., Hori, F., Ishikawa, T., Itoh, M., Izawa, M.,

Kawai, J., Kikuchi, N., Kojima, Y., Matsuyama, T., Niitsuma, H., Oda, H.,

Owa, C., Sato, K., Shibata, Y., Shigemoto, Y., Shiraki, T., Soabe, Y.,

Sugihara, I., Suzuki, H., Suzuki, H., Tateo, M., Tomaru, Y.,

Tomimaga, N., Watanabe, S., Yagame, M., Yamamura, T., Yokota, T.,

Yoshino, M., Muramatsu, M., Okazaki, Y. and Hayashizaki, Y.

RIKEN Mouse ESTs

TITLE Unpublished (1999)

COMMENT

On Mar 10, 1998 this sequence version replaced gi:2948169.

Contact: Chie Owa

Genome Science Laboratory

RIKEN

3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Tel: 81-298-36-9145

Fax: 81-298-36-9098

Email: genome-resetc.riken.go.jp

Thermostabilization and thermoactivation of thermolabile enzymes by trehalose and its application for the synthesis of full length cDNA (Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (<http://genome.rtc.riken.go.jp>) for further details.

FEATURES

Location/Qualifiers

1. .284

/organism="Mus musculus"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="2810052C23"

/clone_lib="Mus musculus C57BL/6J 10-11 day embryo"

/sex="mixed"

/dev_stage="10-11 day embryo"

BASE COUNT 66 a 88 c 65 g 65 t

ORIGIN

alignment_scores:
Quality: 60.50 Length: 61
Ratio: 1.952 Gaps: 2
Percent Similarity: 50.820 Percent Identity: 29.508

alignment_block:

US-09-209-961-20 x AV139690/rev ..

Align seg 1/1 to reverse of: AV139690 from: 1 to: 284

```

4 SerLeuTrpGlyHisLeuGlyValVallylsAlaAsnGlyLeuLeuLe 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
226 CGAGTGTGGGCGAGTGCACAAACATAGTACAGCGCATGGCGTCTC.... 182

20 uGlnThrArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGly 37
: |||||: |||||: |||||: |||||: |||||: |||||: |||||:
181 .....AGCTCGGCA 172

37 etValThrMetVallylsTrpLeuLeuLeuMetThrPheMetAla... 52
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
171 CTGTGTGAATGAAAAACACTGCCCTTGTGGCCACACTTCTGGCAGG 122

53 .....GlyCysArgGlyMetIle 58
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
121 GTGTCCTCCGTCGCGGCGACAGTGGGTCTCTG 89

```

seq_name: gb_est6:N88961

seq_documentation_block:

LOCUS N88961 159 bp mRNA EST 02-APR-1996

DEFINITION K6909f Human fetal heart, Lambda ZAP Express Homo sapiens cDNA

clone K6909 5' similar to EST(YF41H07.R1 CONTAINS MER22 REPETITIVE

ELEMENT), mRNA sequence.

ACCESSION N88961

NID g1442291

VERSION N88961.1 GI:1442291

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 159)

AUTHORS Liew, C.C.

TITLE cDNAs from fetal heart (1996)

JOURNAL
COMMENT

Unpublished (1996)
On Apr 14, 1993 this sequence version replaced gi:638496.

Contact: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
Seq primer: GAAATTAACCTCTCACTAAAGGG.

FEATURES

source

Location/Qualifiers
1. .159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="K6909"
/clone_lib="Human fetal heart, Lambda 2AP Express"
/lab_host="E. coli XL1-Blue"
/note="Vector: Lambda 2AP Express; Site_1: EcoRI; Site_2: XhoI; mRNA was purified from human fetal hearts (8-10 weeks). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into predigested lambda 2AP Express."
57 a 35 c 56 g 11 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 58.00 Length: 35
Ratio: 2.762 Gaps: 2
Percent Similarity: 60.000 Percent Identity: 42.857

alignment_block:

US-09-209-961-20 x N88961/rev ..

Align seg 1/1 to reverse of: N88961 from: 1 to: 159

24 LysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValThrMe 40
||||| : : : : :
100 AAGCCTTATCTTCCTCGCATCCTT.....GGCCTGCTTCTGG 60
40 tValLysLysTrpLeuLeuMetThrPheMetAlaGly.....C 54
: : : : :
59 GCTGTTTCAGTGCTCTCTCTGCCACCTTCGCCGCCGACATGGCGCCT 10
54 ysArg 55
|||||
9 GCCGC 5

seq_name: gb_gss5:AQ183363

seq_documentation_block:

LOCUS AQ183363 394 bp DNA GSS 31-OCT-1998
DEFINITION HS_3142_B2_B10_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3142 Col=20 Row=D, genomic survey sequence.

ACCESSION

NID AQ183363
VERSION Q3580730
KEYWORDS GSS.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 394)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.

TITLE

Construction of a Characterized Clone Resource for Genomic Sequencing

JOURNAL

Unpublished (1998)

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887

Email: jwallace@u.washington.edu
Sequence Tagged Connector
Plate: 3142 row: D column: 20
Class: BAC ends

High quality sequence stop: 394.

FEATURES

source

Location/Qualifiers
1. .394
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Plate=3142 Col=20 Row=D"
/clone_lib="CIT Approved Human Genomic Sperm Library D"
/sex="male"
/note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
117 a 64 c 67 g 146 t

BASE COUNT
ORIGIN

alignment_scores:

Quality: 61.50 Length: 45
Ratio: 2.050 Gaps: 2
Percent Similarity: 56.667 Percent Identity: 37.778

alignment_block:

US-09-209-961-20 x AQ183363/rev ..

Align seg 1/1 to reverse of: AQ183363 from: 1 to: 394

18 LeulleLeuGlnThrArgLysProHisThrGlyAsnHisLeu...LysTh 33
||||| : : : : :
153 CTTTITTTTAACAATGAGTGGCAATGACGGTAACCATATCAAGAAAC 104
33 rSerGlyGlyMetValThrMetVal.....LysLysTrpLeuLeuLeuM 48
: : : : :
103 TCAATGCAGATATATCACTATGTTGTAGAAAAAATATGCTTCTACTGC 54
48 etThrPheMetAlaGlyCysArgGlyMetIleTyr 59
: : : : :
53 TCACAGCTCTTCATATGCTGCTCATGCTCTTTTATAC 19

seq_name: gb_gss4:AQ173665

seq_documentation_block:

LOCUS AQ173665 510 bp DNA GSS 16-OCT-1998
DEFINITION HS_3204_B1_D04_77 CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=3204 Col=7 Row=H, genomic survey sequence.

ACCESSION

NID AQ173665
VERSION Q3571032
KEYWORDS GSS.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 510)

AUTHORS

Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.

TITLE

Construction of a Characterized Clone Resource for Genomic Sequencing

JOURNAL

Unpublished (1998)

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu
 Sequence Tagged Connector
 Plate: 3204 row: H column: 7
 Class: BAC ends
 High quality sequence stop: 510.

FEATURES

source
 1. .510
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate=3204 Col-7 Row=H"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 198 a 85 c 111 g 116 t
 ORIGIN

alignment_scores:
 Quality: 62.50 Length: 52
 Ratio: 2.016 Gaps: 2
 Percent Similarity: 59.615 Percent Identity: 32.692

alignment_block:
 US-09-209-961-20 x AQ173665 ..

Align seg 1/1 to: AQ173665 from: 1 to: 510

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuLeuGlnTh 22
 |||||
 235 TGAATCATCGC.....ATAGAATCGAATGGAATATATCATCGAATGAC 278

22 rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValt 39
 |||||
 279 TCGAATCGAATCAACATCAACGGAATCAACGGAATATATGGAATGGAAT 328

39 hrMetValLysLysTrpLeuLeuMetThrPheMetAlaGlyCysArg 55
 |||||
 329 CG.....AAGAGAATCATCGAATCGAATCGAATCGAATCATCTAAT 369

56 GlyMet 57
 |||||

370 GGAATG 375

seq_name: gb_gss4:AQ177242

seq_documentation_block:
 LOCUS AQ177242 526 bp DNA GSS 16-OCT-1998
 DEFINITION HS_3033_B1_E05_MR CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-3033 Col-9 Row-J, genomic survey sequence.

ACCESSION AQ177242
 NID 93574609
 VERSION AQ177242.1 GI:3574609
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 526)

AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.

TITLE Construction of a Characterized Clone Resource for Genomic Sequencing

JOURNAL Unpublished (1998)

COMMENT
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 EMail: jwallace@u.washington.edu

Sequence Tagged Connector
 Plate: 3033 row: J column: 9
 Class: BAC ends
 High quality sequence stop: 526.
 Location/Qualifiers
 1. .526

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="plate=3033 Col-9 Row=J"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelOBAC11; BAC Clones in E-Coli DH10B"

BASE COUNT 198 a 93 c 118 g 115 t 2 others
 ORIGIN

alignment_scores:
 Quality: 62.50 Length: 52
 Ratio: 2.016 Gaps: 2
 Percent Similarity: 59.615 Percent Identity: 32.692

alignment_block:
 US-09-209-961-20 x AQ177242 ..

Align seg 1/1 to: AQ177242 from: 1 to: 526

6 TrpGlyHisLeuGlyValValLysAlaAsnGlyLeuLeuLeuLeuGlnTh 22
 |||||
 264 TGAATCATCGC.....ATAGAATCGAATGGAATATATCATCGAATGAC 307

22 rArgLysProHisThrGlyAsnHisLeuLysThrSerGlyGlyMetValt 39
 |||||
 308 TCGAATCGAATCAACATCAACGGAATCAACGGAATATATGGAATGGAAT 357

39 hrMetValLysLysTrpLeuLeuMetThrPheMetAlaGlyCysArg 55
 |||||
 358 CG.....AAGAGAATCATCGAATCGAATCGAATCGAATCATCTAAT 398

56 GlyMet 57
 |||||

399 GGAATG 404

seq_name: gb_gss9:AQ626156

seq_documentation_block:
 LOCUS AQ626156 607 bp DNA GSS 16-JUN-1999
 DEFINITION CITBI-El-2657P22.TR CITBI-El Homo sapiens genomic clone 2657P22, genomic survey sequence.

ACCESSION AQ626156
 NID 95088548
 VERSION AQ626156.1 GI:5088548
 KEYWORDS GSS.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 607)

AUTHORS Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.

TITLE Use of BAC End Sequences from CalTech Libraries for Sequence-Ready

JOURNAL Map Building

COMMENT Unpublished (1997)

Other-GSSs: CITBI-El-2657P22.TF

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
 source
 location/Qualifiers
 1..607
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="2657P22"
 /clone_lib="CITBI-El"
 /sex="male"
 /cell_type="sperm"
 /note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
 CalTech Human BAC Library D"
BASE COUNT 226 a 66 c 114 g 201 t
ORIGIN

alignment_scores:
 Quality: 63.00 Length: 30
 Ratio: 3.000 Gaps: 0
 Percent Similarity: 70.000 Percent Identity: 40.000

alignment_block:
US-09-209-961-20 x A0626156 ..

Align seg 1/1 to: A0626156 from: 1 to: 607

30 HisLeuLysThrSerGlyGlyMetValThrMetValLysLysTyrLeuLe 46
 |||||: |||||:|||||: :: ||| ||
70 CATTTAAATGAATGGGGGAGTAGTAAGCTGTTATAGTATTGGAAACT 119

46 uLeuMetThrPheMetAlaGlyCysArgGlyMetIleTyr 59
 |||||:||||| |||||:|||||:|||||
120 ATGGGTAAACATTTAACACAGTGTGTTAACAGTTTATCTAT 159

OM of: US-09-209-961-21 to: GenEmbl:* out_format : pfs
 Date: Dec 27, 1999 2:13 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:
 -MODEL=frameat-p2n.model -DEV=xlp
 -Q/-cgn2_l/USPRO_spool/US09209961/runat_21211999_101043_17910/app_query.fasta.1
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
 -LIST=45 -LOCALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
 -OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY
 -WAIT -THREADS=1

Search information block:

Query: US-09-209-961-21
 Query length: 53
 Database: GenEmbl:*
 Database sequences: 780561
 Database length: 2137953050
 Search time (sec): 1994.760000

score_list:

Sequence	Strd	Orig	Zscore	Escore	Len	Documentation
gb_v1:AF134203	+	273.00	681.70	5.7e-30	436	AF134203 Porcine circovirus PCV
gb_v1:AF135393	+	273.00	681.66	5.7e-30	438	AF135393 Porcine circovirus PCV
gb_v1:AF118097	+	273.00	677.59	9.7e-29	700	AF118097 Porcine circovirus ty
gb_v1:AF055392	+	273.00	669.57	2.7e-29	1768	AF055392 Porcine circovirus ty
gb_v1:AF085695	+	273.00	669.57	2.7e-29	1768	AF085695 Porcine circovirus st
gb_v1:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_v1:AF086835	+	273.00	669.57	2.7e-29	1768	AF086835 Porcine circovirus st
gb_v1:AF086836	+	273.00	669.57	2.7e-29	1768	AF086836 Porcine circovirus st
gb_v1:AF109397	+	273.00	669.57	2.7e-29	1768	AF109397 Bovine circovirus, co
gb_v1:AF112862	+	273.00	669.57	2.7e-29	1768	AF112862 Porcine circovirus, c
gb_v1:AF055391	+	270.00	661.98	7.2e-29	1768	AF055391 Porcine circovirus ty
gb_v1:PCAT3185	+	269.00	659.45	9.9e-29	1768	AF223185 Porcine circovirus DN
gb_v1:AF109399	+	267.00	654.39	1.9e-28	1768	AF109399 Porcine circovirus ty
gb_v1:AF027217	+	266.00	651.86	2.6e-28	1768	AF027217 Porcine circovirus st
gb_v1:AF147751	+	265.00	649.33	3.6e-28	1768	AF147751 Porcine circovirus ty
gb_v1:AF154679	+	262.00	641.74	9.6e-28	1768	AF154679 Porcine circovirus, c
gb_v1:AF166528	+	259.00	634.15	2.5e-27	1768	AF166528 Porcine circovirus, c
gb_v1:AF109398	+	249.00	608.85	6.5e-26	1768	AF109398 Porcine circovirus ty
gb_v1:AF117753	+	240.00	608.85	6.5e-26	1768	AF117753 Porcine circovirus ty
gb_v1:AF055393	+	210.00	510.18	2.0e-20	1767	AF055393 Porcine circovirus ty
gb_v1:AF055394	+	210.00	510.18	2.0e-20	1767	AF055394 Porcine circovirus ty
gb_v1:AF071879	+	72.50	162.34	0.4846	1758	AF071879 Porcine circovirus, c
gb_v1:AF012107	+	70.00	156.01	1.09	1759	AF012107 Porcine circovirus, c
gb_v1:PCOMGEN	+	68.00	150.95	2.09	1759	Y09921 Porcine circovirus com
gb_v1:PC049186	+	68.00	150.95	2.09	1759	Y09186 Porcine circovirus, com
gb_pr1:HUNSATRA	+	67.50	145.14	4.40	2970	L01057 Human (clone trp-6) sat
gb_pr1:D50001S06	+	70.00	144.34	4.87	6760	D50006 Human DNA for alpha-pla
gb_in2:AF079059	+	58.00	139.20	9.42	368	AF079059 Mesobuthus martensii c
gb_sts:HS320XB5	+	57.50	137.59	11.59	383	D35365 H. sapiens (D35365) DNA
gb_in2:AR091022	+	57.00	136.79	12.84	363	AF091022 E. coli phage phi11 gene
gb_p11:YSCITR1	+	64.00	136.34	13.60	2953	D90352 S. cerevisiae ITR1 gene
gb_p12:AF004707	+	57.50	136.25	13.76	447	AF004707 Gigaspora albida stra
gb_v1:AF106408	+	57.00	136.10	14.02	393	AF106408 HIV-1 isolate T2B5041
gb_pr2:HSU85H7	+	72.50	135.93	14.33	37054	I 273900 Human DNA sequence fcd
gb_pr1:AS012770	+	62.00	135.01	16.12	1919	AS012770 Homo sapiens gene for
gb_v1:AF106409	+	56.00	133.64	19.23	390	AF106409 HIV-1 isolate T2B5043
gb_v1:AF106398	+	56.00	133.57	19.40	393	AF106398 HIV-1 isolate T2B5030
gb_v1:AF106452	+	56.00	133.57	19.40	393	AF106452 HIV-1 isolate T2B5092
gb_om:CFDOPR09	+	55.50	133.48	19.61	343	X64970 C. familiaris mRNA DOPCR
gb_ro:RATKINKL	+	59.00	132.38	22.60	1083	M11884 Rat low molecular weigh
gb_ro:AF017085	+	53.50	132.25	22.98	4091	AF017085 Mus musculus BAP-135
gb_in2:AF034556	+	55.50	131.75	24.37	417	AF034556 Oshmarinella rochebru

gb_v1:HIV228318 - 54.50 131.03 26.87 340 ! AJ228318 Human immunodeficie
 gb_v1:AF106390 - 55.00 130.97 27.06 396 ! AF106390 HIV-1 isolate T2B50
 gb_v1:AF106444 - 55.00 130.91 27.29 399 ! AF106444 HIV-1 isolate T2B50

seq_name: gb_v1:AF134203

seq_documentation_block:
 LOCUS AF134203 436 bp ss-DNA VRL 02-JUN-1999

DEFINITION Porcine circovirus PCV-2A amplified sequence.

ACCESSION AF134203

NID 94959580

VERSION AF134203.1 GI:4959580

KEYWORDS

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 436)

AUTHORS Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.

TITLE A PCR assay for detecting and characterizing type-2 circovirus in

plgs, cattle, bison and sheep with various clinical syndromes

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 436)

AUTHORS Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (10-MAR-1999) Virology Laboratory, Manitoba Agriculture,

Veterinary Services, 545 University Crescent, Winnipeg, Manitoba

R3T 5S6, Canada

FEATURES

Location/Qualifiers

1..436

/organism="porcine circovirus"

/strain="PCV-2A"

/db_xref="taxon:46221"

/note="isolated from sheep (ewe)"

1..436

misc-feature

/note="PCR amplification product using oligonucleotide

primers derived from porcine circovirus from pigs

deposited in GenBank Accession Number AF027217"

BASE COUNT 107 a 83 c 103 g 143 t

ORIGIN

alignment_scores:

Quality: 273.00 Length: 53

Ratio: 5.151 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-21 x AF134203

Align seg 1/1 to: AF134203 from: 1 to: 436

1 MetValPheIleIleHisLeuGlyPheIlystrpGlyValPheIlyIlely 17

185 ATGGTGTATTTATTCATTAGGTTTAAAGTGGGGCTTTAAGATTAA 234

17 gPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValy 34

235 ATTCCTCTGAATTGTACATACATGTTTACACGGATATTGATGCTGCTG 284

34 alpHeThrValPheGluArgSerAlaGluAlatyrrValValHisIleSer 50

285 TATTTACTGTTTTCGAACGACAGTCCGAGGCTTACGTGTCACATTCT 334

51 ArgGlyLeu 53

335 AGAGGTTTG 343

seq_name: gb_v1:AF135393

seq_documentation_block:

LOCUS AF135393 438 bp ss-DNA VRL 02-JUN-1999

DEFINITION Porcine circovirus PCV-2A amplified sequence.

ACCESSION AF135393

```

NID          94959706
VERSION      AF135393.1  GI:4959706
KEYWORDS     porcine circovirus.
SOURCE       porcine circovirus.
ORGANISM     Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE    1 (bases 1 to 438)
AUTHORS      Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.
TITLE        A PCR assay for detecting and characterizing type-2 circovirus in
JOURNAL      pigs, cattle, bison and sheep with various clinical syndromes
              Unpublished
REFERENCE    2 (bases 1 to 438)
AUTHORS      Hamel,A.L., Lin,L., Sachvie,C., Grudeski,E. and Nayar,G.P.S.
TITLE        Direct Submission
JOURNAL      Submitted (17-MAR-1999) Virology Laboratory, Veterinary Services,
              Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba
              R3T 5S6, Canada

FEATURES     Location/Qualifiers
              1..438
               /organism="porcine circovirus"
               /strain="PCV-2A"
               /db_xref="taxon:46221"
               /note="isolated from bison"
              1..438
               misc_feature
               /note="PCR amplification product using oligonucleotide
               primers derived from porcine circovirus from pig deposited
               in GenBank Accession Number AF027217"

BASE COUNT   106 a   83 c   104 g   145 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-209-961-21 x AF135393  ..

  Align seg 1/1 to: AF135393 from: 1 to: 438

      1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
      186 ATGGTGTATTATTATTCATTAGGTTAAAGTGGGGGCTTTAAGATTAA 235
      17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValV 34
      236 ATTCTCTGAATTGTACATACATAGTTACACGGATATTGTAGTCTGTCG 285
      34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
      286 TATTTACTGTTTTCGACGACGAGTGGCGAGGCTACGTGGTCCACATTCT 335
      51 ArgGlyLeu 53
      336 AGAGGTTTG 344

  seq_name: gb_vi:AF118097

seq_documentation_block:
  LOCUS      AF118097      700 bp      DNA      VRL      27-JUL-1999
  DEFINITION Porcine circovirus type 2 isolate IAF-4370 unknown gene.
  ACCESSION  AF118097
  NID        95596427
  VERSION    AF118097.1  GI:5596427
  KEYWORDS   porcine circovirus type 2.
  SOURCE     porcine circovirus type 2
  ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 700)
  AUTHORS    Guardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
  TITLE      Multiplex PCR for detection and typing of porcine circoviruses
  JOURNAL    J. Clin. Microbiol. (1999) In press

REFERENCE    2 (bases 1 to 700)
AUTHORS      Guardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
TITLE        Direct Submission
JOURNAL      Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des
              Prairies, Laval, Que H7N 4Z3, Canada

FEATURES     Location/Qualifiers
              1..700
               /organism="porcine circovirus type 2"
               /isolate="IAF-4370"
               /db_xref="taxon:85708"
               /note="virus isolated from lung, lymph node, spleen and
               tonsil tissue from pigs affected by postweaning
               multisystemic wasting syndrome"
               complement(42..>700)
               /note="ORF2"
               /codon_start=3
               /product="unknown"
               /protein_id="AAD45581.1"
               /db_xref="PID:95596428"
               /db_xref="GI:5596428"
               /translation="BSHLGQILRRRRLVPHRRVYWRKNGIFNRLSRTFCGYTKA
               TTVTPSWAVDMRENIDDFVPGGTNKAISIPFEYRIKRYKVAEWPSPVQGBRG
               VGSTAVILDNDFVTATATYDPIVYNSRRHTIPQPFHSRIFTKPYLDSTIDIFQ
               PNKNRQLWMLRLOTGRNVDHVGLGTAFENSKYDQDYNIIRVTWYVQPRENLKDPPLKP
               "

BASE COUNT   156 a   127 c   210 g   207 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-209-961-21 x AF118097  ..

  Align seg 1/1 to: AF118097 from: 1 to: 700

      1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
      24 ATGGTGTATTATTATTCATTAGGTTAAAGTGGGGGCTTTAAGATTAA 73
      17 sPheSerGluLeuTyriIleHisGlyTyThrAspIleValValLeuValV 34
      74 ATTCTCTGAATTGTACATACATAGTTACACGGATATTGTAGTCTGTCG 123
      34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
      124 TATTTACTGTTTTCGACGACGAGTGGCGAGGCTACGTGGTCCACATTCT 173
      51 ArgGlyLeu 53
      174 AGAGGTTTG 182

  seq_name: gb_vi:AF055392

seq_documentation_block:
  LOCUS      AF055392      1768 bp      DNA      circular      VRL      13-SEP-1998
  DEFINITION Porcine circovirus Type II from Canada, complete genome.
  ACCESSION  AF055392
  NID        93598808
  VERSION    AF055392.1  GI:3598808
  KEYWORDS   porcine circovirus.
  SOURCE     porcine circovirus
  ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 1768)
  AUTHORS    Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
              Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
  TITLE      Characterization of novel circovirus DNAs associated with wasting
              syndromes in pigs
  JOURNAL    J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)

```

```
98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Queen's University of, Belfast, Stoney Road, Belfast, Northern Ireland, BT4 3SD, U.K.
FEATURES
    source
        1..1768
            /organism="porcine circovirus"
            /specific_host="Sus scrofa"
            /strain="Porcine circovirus Type II"
            /db_xref="taxon:46221"
            /clone="Imp.1010-Stoon ECORI No.14"
            /country="Canada"
            complement(101..190)
            /codon_start=1
            /product="ORF6"
            /protein_id="AAC35319.1"
            /db_xref="PID:g3598819"
            /db_xref="GI:3598819"
            /translation="MASTPASPAPSDILSSVPQSQRPPQWWT"
            103..210
            /codon_start=1
            /product="ORF10"
            /protein_id="AAC35317.1"
            /db_xref="PID:g3598817"
            /db_xref="GI:3598817"
            /translation="MSTAQEGVVTVVRLTVYPKVRRRLKMPFFLLQR"
            261..320
            /codon_start=1
            /product="ORF7"
            /protein_id="AAC35318.1"
            /db_xref="PID:g3598818"
            /db_xref="GI:3598818"
            /translation="MAAGAVSSSAVTPPWIRHS"
            complement(311..439)
            /note="predicted 4.6 kDa protein"
            /codon_start=1
            /product="ORF9"
            /protein_id="AAC35315.1"
            /db_xref="PID:g3598815"
            /db_xref="GI:3598815"
            /translation="MVLGSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRFQL"
            390..1342
            /codon_start=1
            /product="putative Rep protein (ORF1)"
            /protein_id="AAC35309.1"
            /db_xref="PID:g3598809"
            /db_xref="GI:3598809"
            /translation="MPSKNGRSGPQPHKRWFTLNNPSEDEKKIRELPISLFDYFI
            VGEENEGRTFHLQGFANFKOTFNKRWYLGARCHEIAKAGTDQONKEYCKEKN
            LITCCAPRSQORSLSYAVSTLLESGSLTVAEQHPVTFVRNFRGLAELLKVSQGM
            OKRDWNTNHHVIVGPPCGCKSKWAANFADPTTYMKPPRNKWDGYPHGEVVDIDFY
            GHLPWDLRLCLRDYELTVETKGTGTFPFLARSILITSNQPLEWYSTAVPAVEALYR
            RITSLVFNKATLQSTQSGQFVTLSPPCPEFFPEYNI"
            674..679
            complement(704..1018)
            /note="predicted 11.9 kDa protein"
            /codon_start=1
            /product="ORF3"
            /protein_id="AAC35311.1"
            /db_xref="PID:g3598811"
            /db_xref="GI:3598811"
            /translation="MTIPLVSRWFPVCGFRVCKYSSPFAFTTPRPHNDVVIGLPI
            TLLHFAHQKFSQPAEISDKRYRVLNCGHQTPALQQGTHSSRQVTPLSLRSSSTF
            NK"
            complement(733..912)
            /note="predicted 6.5 kDa protein"
            /codon_start=1
            /product="ORF4"
            /protein_id="AAC35312.1"
            /db_xref="PID:g3598812"

polyA_signal
CDS
    1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIle 17
    1363 ATGGTATTTTATTATTATTTAGGTTTAAAGTGGGGGTCTTTAAGATTAA 1412
    17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
    1413 ATTCCTCTGAATGTACATACATGTTACACGGATATTAGTCCTGGTCG 1462
    34 alphaThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
    1463 TATTACTGTTTTCGAACGACGCGCGAGCGCTACGTGGTCCACATTCC 1512
    51 ArgGlyLeu 53
    1513 AGAGGTTTG 1521
    seq_name: gb_v1:AF085695

/db_xref="GI:3598812"
/translation="MTCTLVQSRFCIFPLTFKSSASPRKELTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35314.1"
/db_xref="PID:g3598814"
/db_xref="GI:3598814"
/translation="MDIDHTVSVDPHTAASHKSHO"
1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35316.1"
/db_xref="PID:g3598816"
/db_xref="GI:3598816"
/translation="MNNKNHYEVIKKTO"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35313.1"
/db_xref="PID:g3598813"
/db_xref="GI:3598813"
/translation="MVFIIHLGFKWGVKIKFSELYIHGYTDIVVLVVTVPFERSAEA
YVHISRGL"
complement(1369..1374)
complement(join(1381..1768,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:g3598810"
/db_xref="GI:3598810"
/translation="MTYPRRYRRRRHRRPSHLGQLRRRLVHPRIHYRWRRKNGI
FNTLSRTFGYTVKRTTTPTSWAVDMFRKIDDFVPPGGTNKISIPFYYIRKVK
VEFWPCSPITQDGRGVSTAVILDNDFVTATLYDPVYNSRHTIPQPFYSHSRY
FTPKPVLDSIDYDFQPNKNRQLWLRLQTSNVHVLGAAAFENSKYDQYINIRVTM
VQRFENLKDPPPLKP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores
    Quality: 273.00 Length: 53
    Ratio: 5.151 Gaps: 0
    Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block
    US-09-209-961-21 x AF055392 ..
Align seg 1/1 to: AF055392 from: 1 to: 1768
    1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIle 17
    1363 ATGGTATTTTATTATTATTTAGGTTTAAAGTGGGGGTCTTTAAGATTAA 1412
    17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
    1413 ATTCCTCTGAATGTACATACATGTTACACGGATATTAGTCCTGGTCG 1462
    34 alphaThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
    1463 TATTACTGTTTTCGAACGACGCGAGCGCTACGTGGTCCACATTCC 1512
    51 ArgGlyLeu 53
    1513 AGAGGTTTG 1521
    seq_name: gb_v1:AF085695
```

```
seq_documentation_block:
LOCUS       AF085695      1768 bp      DNA      circular      VRL      30-SEP-1998
DEFINITION   Porcine circovirus strain 412, complete genome.
ACCESSION   AF085695
NID         93668362
VERSION     AF085695.1  GI:3668362
KEYWORDS    porcine circovirus.
SOURCE      porcine circovirus.
ORGANISM    porcine circovirus.
REFERENCE   1 (bases 1 to 1768)
AUTHORS     Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE       Emergence of a new porcine circovirus.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1768)
AUTHORS     Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE       Direct Submission
JOURNAL     Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,
            University of Saskatchewan, 120 Veterinary Road, Saskatoon,
            Saskatchewan S7N 5E3, Canada
FEATURES             Location/Qualifiers
     source          1..1768
                     /organism="porcine circovirus"
                     /strain="412"
                     /db_xref="taxon:46221"
     CDS             51..995
                     /codon_start=1
                     /product="p35.8"
                     /protein_id="AAC61860.1"
                     /db_xref="PID:G3668363"
                     /db_xref="GI:3668363"
                     /translation="WPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
                     VGEENEGRTHPLGFANFYKQFENKVKWLGARCHEKAKGDDQNKKEVCSREGN
                     LIECGAPQRSDLSAVSTLIESGLITVAKQHPVTFVKNFRGLAEELKLVSGKM
                     OKRDKNTVHFIVGPPGCKSKWAFANPETTYKPPKNKWDGYHGEKVVIDDFY
                     GWPWDLLRLCDRLPLVTKTGTVFPLARSILITSNQPLEWYSSTAAPVAEALYR
                     RTLSLVFNKNATQKSTEGGQFVTLSPPCPEPPEYIN"
                     complement(357..671)
                     /codon_start=1
                     /product="p12.1"
                     /protein_id="AAC61865.1"
                     /db_xref="PID:G3668368"
                     /db_xref="GI:3668368"
                     /translation="WVTIPPLYRFPVCGFRVCKISSPFAETTPRPHNEVYIGFPI
                     TLHHPFAHFQFSQPAELFDKIRVLLCNGHONPALQOQTHSRQVTPLSLRSSTF
                     NK"
     CDS             complement(386..565)
                     /codon_start=1
                     /product="p6.6"
                     /protein_id="AAC61866.1"
                     /db_xref="PID:G3668369"
                     /db_xref="GI:3668369"
                     /translation="MKCTLVFSQSFCEIFPLTFKSSASPRKFLTNVTGCCFAIVTRIPL
                     SNKVLTAVDRLRCP"
                     553..732
                     /codon_start=1
                     /product="p6.6"
                     /protein_id="AAC61861.1"
                     /db_xref="PID:G3668364"
                     /db_xref="GI:3668364"
                     /translation="WYTSLWGLHVGVKANGLLILOTRPHTGNHLKTSGMVTVKKW
                     LLMTFMAGCRNIY"
     CDS             833..970
                     /codon_start=1
                     /product="p5.0"
                     /protein_id="AAC61862.1"
                     /db_xref="PID:G3668365"
                     /db_xref="GI:3668365"
                     /translation="MVLLNCPSCRSLSSEDFYLGILECYKTIHGGRGVRHPFPPM
                     P"
     CDS             1016..1177
                     /codon_start=1
```

```

/product="p6.2"
/protein_id="AAC61863.1"
/db_xref="PID:G3668366"
/db_xref="GI:3668366"
/translation="MVFIHLGFKWGVFKIKFSELYIHGYTDIWLWVTFVFSABEA
YVHISRGL"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61864.1"
/db_xref="PID:G3668367"
/db_xref="GI:3668367"
/translation="MTYIPRRYRRRRPRSHLQILRRPWLVPHRHYRWRKNGI
FNTRLSRTFGYTVKRTVTTPSWAVDMRFKIDFVPPGGTGNKISIPFVYIRKVK
VEFWPCSPITQDGRGVSTAVLLDDNFVKATALTVDYPVYNSRHTIPQPFVSHSY
FTPKPVLDSTIDYFQPNKRNQLWLRLQISGNVDHVLGTAFENSKYDQDYNIRVTHY
VQREFNLKDPPLP"
BASE COUNT      463 a      362 c      481 g      462 t
ORIGIN
alignment_scores:
    Quality: 273.00      Length: 53
    Ratio: 5.151      Gaps: 0
    Percent Similarity: 100.000      Percent Identity: 100.000
alignment_block:
US-09-209-961-21 x AF085695
Align seg 1/1 to: AF085695 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTTTATATTCATTAGGTTCAAGTGGGGGCTTTAGATAA 1065
17 sPheSerGluLeuTyrlleHisGlyTyThrAspileValValLeuValV 34
1066 ATTCTCTGAATTGACATACATGTTACACGATATTGTAGTCCTGGTCG 1115
34 alpHeThrValPheGluArgSerAlaGluAlaTyrrValValHisIleSer 50
1116 TATTACTGTTTTCAGCAGTCGCGAGGCTACGTGTCGCACATTTC 1165
51 ArgGlyLeu 53
1166 AGAGTTTG 1174
seq_name: gb_vi:AF086834
seq_documentation_block:
LOCUS       AF086834      1768 bp      DNA      circular      VRL      29-SEP-1998
DEFINITION   Porcine circovirus strain B9, complete genome.
ACCESSION   AF086834
NID         93661515
VERSION     AF086834.1  GI:3661515
KEYWORDS    porcine circovirus.
SOURCE      porcine circovirus.
ORGANISM    porcine circovirus.
REFERENCE   1 (bases 1 to 1768)
AUTHORS     Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE       Emergence of a new porcine circovirus
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1768)
AUTHORS     Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE       Direct Submission
JOURNAL     Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
            Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES             Location/Qualifiers
     source          1..1768
                     /organism="porcine circovirus"
                     /strain="B9"
                     /db_xref="taxon:46221"
```

```
CDS          51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:g3661516"
/db_xref="GI:3661516"
/translat="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPLHOGFANFVKOTFNKVKWYLGARCHIEKAKGTDOONKEVCSKEGN
LLECGAPRSQGRSDLSAVSTLESGLVTAKQHPVTFVKNFGLAELLKVSCKM
QRDKNTNHFIVGPGCGSKWAANFANPETTYWPKPKNWGDYHGEKVVYIDDFY
GMLPDDLLRLCDRYPLTVTKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFYEINY"
complement(1034..1735)
/codon_start=1
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:g3661517"
/db_xref="GI:3661517"
/translat="MTYPRRRYRRRRPRSHLQILRRRPWLVPVPHRYRWRRKNGI
FNRLSRTGYTVKATVTTPSWAVDMRENIIDFVPGGGTGNKISIPFEYVIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVTKATALTYPYVNSRHTIPQPFYSHSRY
FTPKPVDLSDIDFQPNKRNQLWLRLQTSRNVHDVHGLGTAFENSKYDQDYNIRVTM
VQFREFNLKDPPLKP"
BASE COUNT  463 a 357 c 482 g 466 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF086834 ..
Align seg 1/1 to: AF086834 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTGTATTTATTTATTTAGGTTTAAAGTGGGGTCTTTAAAGATTAA 1065

17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValIleuVal 34
1066 ATTCCTGAATTGACATACATAGTTACACGGAATTTAGTCTCGGTGCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTACTGTTTTCGACGACGAGTCCGAGGCTACGTGTCACATTTCT 1165

51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF086835

seq_documentation_block:
LOCUS      AF086835      1768 bp      DNA      circular      VRL
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION  AF086835
NID        93661518
VERSION    AF086835.1 GI:3661518
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE      Emergence of a new porcine circovirus
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE      Direct Submission
JOURNAL    Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
```

```
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
Location/Qualifiers
1..1768
/organism="porcine circovirus"
/db_xref="9741"
/db_xref="taxon:46221"
51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61739.1"
/db_xref="PID:g3661519"
/db_xref="GI:3661519"
/translat="MPSKNGRSGPQPHKRWFTLNNPSENKRRKKIRELPISLFDYFI
VGEENEGRTPLHOGFANFVKOTFNKVKWYLGARCHIEKAKGTDOONKEVCSKEGN
LLECGAPRSQGRSDLSAVSTLESGLVTAKQHPVTFVKNFGLAELLKVSCKM
QRDKNTNHFIVGPGCGSKWAANFANPETTYWPKPKNWGDYHGEKVVYIDDFY
GMLPDDLLRLCDRYPLTVTKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFYEINY"
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61740.1"
/db_xref="PID:g3661520"
/db_xref="GI:3661520"
/translat="MTYPRRRYRRRRPRSHLQILRRRPWLVPVPHRYRWRRKNGI
FNRLSRTGYTVKATVTTPSWAVDMRENIIDFVPGGGTGNKISIPFEYVIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVTKATALTYPYVNSRHTIPQPFYSHSRY
FTPKPVDLSDIDFQPNKRNQLWLRLQTSRNVHDVHGLGTAFENSKYDQDYNIRVTM
VQFREFNLKDPPLKP"
BASE COUNT  464 a 360 c 480 g 464 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF086835 ..
Align seg 1/1 to: AF086835 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTGTATTTATTTATTTAGGTTTAAAGTGGGGTCTTTAAAGATTAA 1065

17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValIleuVal 34
1066 ATTCCTGAATTGACATACATAGTTACACGGAATTTAGTCTCGGTGCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTACTGTTTTCGACGACGAGTCCGAGGCTACGTGTCACATTTCT 1165

51 ArgGlyLeu 53
1166 AGAGGTTTG 1174

seq_name: gb_vi:AF086836

seq_documentation_block:
LOCUS      AF086836      1768 bp      DNA      circular      VRL
DEFINITION Porcine circovirus strain M236, complete genome.
ACCESSION  AF086836
NID        93661521
VERSION    AF086836.1 GI:3661521
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
```

```

TITLE      Emergence of a new porcine circovirus
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE      Direct Submission
JOURNAL    Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
            Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES   Location/Qualifiers
            source
            1..1768
            /organism="porcine circovirus"
            /strain="M226"
            /db_xref="taxon:46221"
            CDS
            51..995
            /product="p35.8"
            /protein_id="AAC61741.1"
            /db_xref="PID:G3661522"
            /db_xref="GI:3661522"
            /translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
            VGEENEGRTPLHGFANFVKQTFNKVYFGARHIEKAKGTDOONKEKVCSEGN
            LLIEGAPESQSGSDLSAVSTLLESGLITVAQHPVTEVKNFERGLAEKLVKSGKM
            OKRDKTNVHVIYVGGPCGSKWAANFADPETTYKPKNKWWDGYHGEVYVDDFY
            GWLPWDDLRLCDRYPLTVETKGTGVTPLARSILITSNQTPLEWYSSTAVPAVEALYR
            RITSLVFWKNATKQSTEEGGQFVTLSPCCPEFFEYNY"
            complement(1034..1735)
            CDS
            /codon_start=1
            /product="p27.9"
            /protein_id="AAC61742.1"
            /db_xref="PID:G3661523"
            /db_xref="GI:3661523"
            /translation="MTYPRRRYRRRRRPSHLQILRRRPLVHPHRYRWRKNGKI
            FNRSLRFGYVTKRTVTTPSWAVDMREKIDDFVPGGNTKISIPPEYRIRKVK
            VEFWPCSPITQDRCVGSTAVILDNENVTAKATAYDPVYNYSSRHTIPOPFSHSRY
            FTKPKVLSTIDYDFPNKRNQLMLRLQTSNGNVHVLGTATFENSKYDQDYNIRVTMI
            VQREFNLKDPLEP"
BASE COUNT 463 a 363 c 480 g 462 t
ORIGIN

alignment_scores:
  Quality: 273.00      Length: 53
  Ratio: 5.151         Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF086836 ..
Align seg 1/1 to: AF086836 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheIleYsrPGlyValPheIleYleY 17
|||||
1016 ATGCTTTTATTATCTATTAGGTTCAAGTGGGGGTCCTTAAGATTAA 1065
|||||
17 sPheSerGluLeuTyrtIleHisGlyTyrtThrAspIleValValLeuValV 34
|||||
1066 ATTCTCTGAATTGACATACATGCTTACCGGATATGCTAGCTCTGGTCG 1115
|||||
34 alPheThrValPheGluArgSerAlaGluAlaTyrtValValHisIleSer 50
|||||
1116 TATTACTGTTTTCGACGAGTCCGAGGCTACGTGGTCCACATTTC 1165
|||||
51 ArgGlyLeu 53
|||||
1166 AGAGGTTTG 1174

seq_name: gb_v1:AF109397

seq_documentation_block:
LOCUS      AF109397 1768 bp DNA circular VRL
DEFINITION Bovine circovirus, complete genome.
ACCESSION  AF109397
NID        94210313
VERSION    AF109397.1 GI:4210313

```

```

KEYWORDS   bovine circovirus.
SOURCE     bovine circovirus
ORGANISM   Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Hamel,A.L. and Nayar,G.P.S.
TITLE      Nucleotide sequence of a circovirus detected in cattle with various
            clinical syndromes
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Hamel,A.L. and Nayar,G.P.S.
TITLE      Direct Submission
JOURNAL    Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
            Laboratory, Manitoba Agriculture, 545 University Crescent,
            Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE  3 (bases 1 to 1768)
AUTHORS    Hamel,A.L. and Nayar,G.P.S.
TITLE      Direct Submission
JOURNAL    Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
            Laboratory, Manitoba Agriculture, 545 University Crescent,
            Winnipeg, Manitoba R3T 5S6, Canada
REMARK     Sequence update by submitter
COMMENT     On Feb 2, 1999 this sequence version replaced gi:4028609.
FEATURES   Location/Qualifiers
            1..1768
            /organism="bovine circovirus"
            /viroion
            /db_xref="taxon:85542"
            /note="BCV
            sequence was obtained from several overlapping PCRs using
            DNA extracted from various tissues from cattle; similar to
            Porcine circovirus sequence presented in GenBank Accession
            Number AF027217"
            repeat_region
            13..35
            /rpt_type=tandem
            /rpt_unit=13..18
            51..995
            /note="ORF-1: similar to Porcine circovirus ORF1 encoded
            by the sequence presented in GenBank Accession Number
            AF027217."
            /codon_start=1
            /product="putative Rep and coat protein"
            /protein_id="AAD11928.1"
            /db_xref="PID:g4210314"
            /db_xref="GI:4210314"
            /translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
            VGEENEGRTPLHGFANFVKQTFNKVYLGARHIEKAKGTDOONKEKVCSEGN
            LLIEGAPESQSGSDLSAVSTLLESGLITVAEQHPVTFVNRFGRLAEKLVKSGKM
            OKRDKTNVHVIYVGGPCGSKWAANFADPETTYKPKNKWWDGYHGEVYVDDFY
            GWLPWDDLRLCDRYPLTVETKGTGVTPLARSILITSNQTPLEWYSSTAVPAVEALYR
            RITSLVFWKNATKQSTEEGGQFVTLSPCCPEFFEYNY"
            complement(357..671)
            /note="similar to Porcine circovirus ORF-3 encoded by the
            sequence presented in GenBank Accession Number AF027217."
            /codon_start=1
            /product="ORF-3"
            /protein_id="AAD11930.1"
            /db_xref="PID:g4210315"
            /db_xref="GI:4210315"
            /translation="MTYIPLVSRWFVCGFRVCKISSPEAFTHPRPHNDVYICLPI
            TLLHFFAHFQKFSQPAEISDKRYRVLNCGHQTALQGGTHSSRQVTPLSRSRSTF
            NK"
            complement(386..565)
            /note="similar to Porcine circovirus ORF-4 encoded by the
            sequence presented in GenBank Accession Number AF027217."
            /codon_start=1
            /product="ORF-4"
            /protein_id="AAD11931.1"
            /db_xref="PID:g4210316"
            /db_xref="GI:4210316"
            /translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCGCSATVTRLPL
            SNKLTAVDRSLRCP"

```

```
CDS
533..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:4028610"
/db_xref="GI:4028610"
/translation="MYTSLWHLGVVVKANGLLILOTRKPHGTGNHLETSGGMVTMVKK
LLLMFMAGCRGMV"
/complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:4028618"
/db_xref="GI:4028618"
/translation="MDIDHTVSVDHPTAASHKSHQ"
983..988
/complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:4028620"
/db_xref="GI:4028620"
/translation="MNNKNHYEVIKTKQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:4028615"
/db_xref="GI:4028615"
/translation="MVFIHLGFKWGVKIFKFSLYIHGVTDIWLWVTFVFSRSEA
YVYHISRLG"
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:4028612"
/db_xref="GI:4028612"
/translation="MTYPRRRYRRRTPRSHLQILRRPMLVHPHRYRWRKNGI
ENRLSRGTGYVTKVTPSWAVDMRENIDDFVPGGNTKISIPFEYRIRVK
VEFWPCSPITQDGVGSFVAILDONFVKATALTVDYVNTSSRTIQQPFSYHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:4028616"
/db_xref="GI:4028616"
/translation="MASSTPASPSDILSLRPLQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:4028619"
/db_xref="GI:4028619"
/translation="MSTAQEGVLTVVALVTPKVRERRVLKMPFFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:4028617"
/db_xref="GI:4028617"
/translation="NAAGAGPSSAVTPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..12)
/note="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
452 a 362 c 493 g 461 t
BASE COUNT
ORIGIN
```

```
alignment_scores:
Quality: 273.00 Length: 53
Ratio: 5.151 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AF109397 ..
Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIlely 17
1016 ATGGTTTATTATTATTAGGTTTAAGTGGGGGCTTTAAGATTAA 1065
17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValV 34
1066 ATTCTCTGAATTGTATACATACGTTACACGGATATTGTAGTCTGGTCG 1115
34 alphaThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
1116 TATTACHTGTTTTCGACGACGTGCGGAGGCTACGTGTCACACATTCT 1165
51 ArgGlyLeu 53
1166 AGAGGTTTG 1174
seq_name: gb_vi:AF112862

seq_documentation_block:
LOCUS AF112862 1768 bp DNA Circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM
porcine circovirus type 2-B.
porcine circovirus type 2-B.
Viruses; SSNA viruses; Circoviridae; Circovirus;
REFERENCE
1 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Nucleotide sequence of four different isolates of porcine
circovirus detected in pigs with various clinical syndromes
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Direct Submission
TITLE
Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
JOURNAL
FEATURES
source
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/note="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/note="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:94106918"
/db_xref="GI:4106918"
/translation="MPSRKNGSGQPQPKRWFTLNNSSEDERKKIRELPISLFDYFI
VGEENEEGRTPHLQGFANFKVTKQFNKYLKARCHIEKAKGTDQNKKEVCSEKN
LLIECCAPRSQGSRLSTAVSLLESGLTVAEQHPVTFYRFRGLAELLKYSKGM
QRDKWKNVHVIVGPPCGSKSAANFADPETTYKPKRKNWDGYHGEVVVDDFY
GWLPWDDLRLCDRIYPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
```



```
polyA_signal
CDS
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEFFPEYIN"
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03087.1"
/db_xref="PID:g4106919"
/db_xref="GI:4106919"
/translation="MVIIPPLVSRWFPVCGFRVCKKISPPFAFTPRPHNDVDIGLPI
TLHFFPAHFQFSQAEISDRKRYRLLCNGHQTALQOGTHSSROVTPLSLRSRSTF
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03088.1"
/db_xref="PID:g4106920"
/db_xref="GI:4106920"
/translation="MTWTLVFSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRSLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03089.1"
/db_xref="PID:g4106921"
/db_xref="GI:4106921"
/translation="MSTSLWGLGVVKGANGLLIQTRKPHGTGNHLETSGGMVTVYKKW
LLMTFMAGCRGMV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03094.1"
/db_xref="PID:g4106926"
/db_xref="GI:4106926"
/translation="MDIDHTVSDVHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03090.1"
/db_xref="PID:g4106922"
/db_xref="GI:4106923"
/translation="MNNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD03091.1"
/db_xref="PID:g4106923"
/db_xref="GI:4106923"
/translation="MVFIIHLGFKWGVKIFSELYIHGYTDIVLVVTVFERSAEA
YVVIISRGL"
complement(1022..1027)
complement(1256..1735)
/translation="MTYIPRRYRRRRHRRSHLGOILRRRPLWHPHRYRWRKNGI
ENTLSRTFGYTKATVTPSWAVDMRNFIDFVPGCGTKNLISVFEYIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALYDIPVYSSRHTIPQPFYSHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03084.1"
/db_xref="PID:g4106916"
/db_xref="GI:4106916"
/translation="MASSTPASPADSILRSLPQSRPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03092.1"
/translation="MSTAQEGVTVVVALTVYKVRERRVLMKMPFFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03093.1"
/db_xref="PID:g4106925"
/db_xref="GI:4106925"
/translation="MAAGAVSSSAVTPPPWIRHS"
join(1750..1768,1..13)
/translation="putative replication site"
join(1762..1768,1..2)
/translation="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186"
448 a 360 c 493 g 467 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 273.00 Length: 53
Ratio: 5.151 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-21 x AFL112862
..
Align seg 1/1 to: AFL112862 from: 1 to: 1768
1 MeVAlpHeIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
|||||
1016 ATGCTTTTATTATTACATACATGTTACGGATATTGTTAGTCTGGTCG 1115
17 sPheSerGlLeuTyriIleHisGlyTyTrAspIleValIleValVal 34
|||||
1066 ATTCTCGAATTGTACATACATGTTACGGATATTGTTAGTCTGGTCG 1115
34 alpHeThrValPheGluArgSerAlaGluIaTyValValHisIleSer 50
|||||
1116 TATTACTGTTTTCGAACGAGTCCGAGGCTACGTGGTCCACATTCT 1165
51 ArgGlyLeu 53
|||||
1166 AGAGGTTG 1174
seq_name: gb_v1:AF055391

seq_documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus type II from USA, complete genome.
ACCESSION AF055391
NID g3598796
VERSION AF055391.1 GI:3598796
KEYWORDS
SOURCE
porcine circovirus.
ORGANISM
porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
JOURNAL
MEDLINE
98418498
REFERENCE
2 (bases 1 to 1768)
Meehan,B.M.
Direct Submission
Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stonely Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1..1768
```



```
ORGANISM      porcine circovirus
REFERENCE     Viruses: ssDNA viruses; Circoviridae; Circovirus.
AUTHORS      1 (bases 1 to 1768)
TITLE        Direct Submission
JOURNAL      Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
              Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
              50011, U.S.A.
REFERENCE     2 (bases 1 to 1768)
AUTHORS      Morozov, I., Sirdinamitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
              Yoon, K.J. and Paul, P.S.
TITLE        Detection of a novel strain of Porcine circovirus in pigs with
              Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL      Unpublished
FEATURES     Location/Qualifiers
              source
              1..1768
              /organism="porcine circovirus"
              /virion
              /isolate="ISU-31"
              /specific_host="Sus scrofa"
              /db_xref="taxon:46221"
              51..995
              /note="ORF1"
              /codon_start=1
              /protein_id="CA11157.1"
              /db_xref="PID:e1310034"
              /db_xref="PID:g3293041"
              /db_xref="GI:3293041"
              /translation="MPSKNGRSGQPQHKRWFTLNPSEDERKKIRELPISLFDYFI
              VGEENEGRTPHLOGFANFKVKTGNKRWYLGARCHEKAKGTDOQNKKEYCSKEGN
              LLIECGAPRSQGRSDLSSTAVSLLESGLVTAEOHPVTFVFNFRGLAELLKVSCKM
              QKRWKTNVHVIVPGGCKSKWAANFADPTTYWKPKNKWDGYPHGEVWVDDFY
              GWLPWDDLRLCDRPLVETKGTVPFLASLITSNQTPLEWYSSTAVPAVEALYR
              RITSLVFWKNATEQSTEGGQFVLSPPCPPEFYINY"
BASE COUNT   452 a 361 c 492 g 463 t
ORIGIN

alignment_scores:
  Quality: 269.00      Length: 53
  Ratio: 5.075         Gaps: 0
  Percent Similarity: 100.000 Percent Identity: 98.113

alignment_block:
US-09-209-961-21 x PCAJ3185 ..
Align seg 1/1 to: PCAJ3185 from: 1 to: 1768
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
:::|||||
1016 ATTGTTTATTATTATTATTATTAGGTTTAACTGGGGGCTTTTAAATAA 1065
17 sPheSerGluLeuTyriIleHisGlyTyriThrasPilleValIleuValy 34
|||||
1066 ATTCTCTGAATTCATACATACATGTTACACGGATATTGTAGTCTGTGCG 1115
34 alPheThrValPheGluArgSerAlaGluIalaTyriValValHisIleSer 50
|||||
1116 TATTACTGTTTTCGAACGACGATGCGGAGGCTACGTGTCACATTCT 1165
51 ArgGlyLeu 53
|||||
1166 AGAGGTTTG 1174
seq_name: gb_vi:AF109399
seq_documentation_block: 1768 bp DNA circular VRL
LOCUS AF109399
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
```

```

/db_xref="GI:4106904"
/translation="MDIDHTVSDVDPHTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEYIKKTQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFIIHLGFKWGVFKIFSELYIHGYTDIVLVVFTVFERSAEA
YVHIISGLSHPOLIPFVILWLEIVINSIKRNFCEVPGVVGELGDCMAGGVVYVG
IGLGGLYYKVI"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MTIPRRRRRRRRHRRSHLQILRRRRLVHPHRYRWKRRKNGI
FNARLSTFYQKATVSTPSWADVNLRFNLDDFPPGGTNRKISIPFYIRLRYK
VFWPCSPITQGRGVGSSAIIILDNEVIRKATAOTYPPVYNSRRHPIPOPFVSHRY
FTPKPVLDSITDVFQPNKRNQMLMRLQTSRNVHDVGLGTAENSKYDQDYNIRVTY
VQREFNLKQPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVVATVYPKVRRRALKMPFLFQR"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRLPQSRPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSSAETPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
Quality: 267.00 Length: 53
Ratio: 5.135 Caps: 0
Percent Similarity: 98.113 Percent Identity: 98.113

alignment_block:
US-09-209-961-21 x AF109399 ..
Align seg 1/1 to: AF109399 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIleY 17
|||||
1016 ATGGTTTTATTATTATAGGTTTAAAGTGGGGTCTTTAAGATTAA 1065
|||||
17 sPheSerGluLeuTyIleHisGlyTyThrAspIleValValLeuValV 34
|||||
1066 ATTCTCTGAATTGTACATACATGTTTACACGGATATTGTAGTCTGTCG 1115
|||||
34 aIPheThrValPheGluArgSerAlaGluAlaTyValValHisIleSer 50
|||||
1116 TATTACTGTTTTCGACGACGAGTCCGAGGCCCTACGTGGTCCACATTCT 1165
|||||
51 ArgGlyLeu 53
|||||
1166 ACTGGTTTG 1174

seq_name: gb_v1:AF027217

seq_documentation_block:
LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998
DEFINITION Porcine circovirus strain pmws PCV, complete genome.
ACCESSION AF027217
NID 92689645
VERSION AF027217.1 GI:2689645
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus.
REFERENCE 1 (bases 1 to 1768)
Viruses; ssDNA viruses; Circoviridae; Circovirus.
AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.
TITLE Nucleotide sequence of porcine circovirus associated with
postweaning multisystemic wasting syndrome in pigs
J. Virol. 72 (6), 5262-5267 (1998)
JOURNAL 98241772
MEDLINE
REFERENCE 2 (bases 1 to 1768)
Hamel,A.L., Lin,L.L. and Nayar,G.P.S.
AUTHORS Direct Submission
TITLE Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services
Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,
Manitoba R3T 5S6, Canada
JOURNAL Location/Qualifiers
FEATURES
source
1..1768
/organism="porcine circovirus"
/strain="pmws PCV"
/db_xref="taxon:46221"
/notes="both strands of seven overlapping PCR fragments
were sequenced; virus isolated from lung, lymph node,
spleen and tonsil tissue from pigs affected by post
weaning multisystemic wasting syndrome"
51..995
/notes="ORF1; similar to Rep protein encoded by
non-pathogenic PCV, GenBank Accession Number U49186;
predicted 35.8 kDa protein"
/codon_start=1
/product="putative Rep protein"
/protein_id="AAC59462.1"
/db_xref="PID:g2689646"
/db_xref="GI:2689646"
/translation="MPSKNGRSGPQHPKRWFTLNPSDEKRRKIRELPISLFDYFI
VGEENEGRTPHLOGFANFVKQTFNKVKYLGARCYIEKAKGTDDQNKKEYSKEGN
LLIECGAPRSGQRSDLSVSTLLESGSLVTAEQHPVTFRNFRGLAELLKVSCKM
OKRDMKTNVHIVGPGCGSKWAANFADPETTKPNRKNWPGYHGEVAVVDDFY
GWLFDLRLCDDRYPLTVETKGTVPFLARSILITSPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVLSPPCPPEPEINY"
117..125
/notes="glycosylation site"
complement(357..671)
/notes="ORF3; predicted 11.9 kDa protein"
/codon_start=1
/protein_id="AAC59464.1"
/db_xref="PID:g2689648"
/db_xref="GI:2689648"
/misc_feature
polyA_signal
CDS

```



```
/specific_host="Sus scrofa"
/db_xref="taxon:85708"
/note="isolated from pigs with postweaning multisystemic
wasting syndrome"
51..995
CDS
/note="ORF1; 35-36kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="PID:g5007011"
/db_xref="GI:5007011"
/translation="MPSKNGRSGPQHRKRWFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRTPLQGFANFVKQTFNKVWILGARCHIEKAKGTQQNKKEYCSKEGN
LLIECAPSQSQSLSTAVSTLSESGSLVTVAEQHPVTFVFNFRGLAELLKVSQKM
QKRDKNVHVIVGPGCGCKSKWAANFADPTTYMKPPRNKWDGYHGEVVVIDDFY
GHLRDDLLRLGDRPFLVETKGTVPFLFARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATEQSTEEGGQFVTLSPPCPEFFPEINY"
complement(1034..1735)
/note="ORF2; 27-28kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="PID:g5007012"
/db_xref="GI:5007012"
/translation="MTYPRRRYRRRRHRRSHLQILRRRPWLVPVHRRYRRKNGI
FNRLSRTFGYATVKTPTPSWAVDMRFNIDDFVPGGNTKISIPPEYVIRKVK
VEFWCSPYTGDRGVGSTAVILDDNFVKATALTYDPVYNYSRHTIQQPFSCHSRY
FTPKPVLDSITIDFFPNRRNQLWRLQTSRNVDHVLGLTAFENSIYDQINIRVTMI
VQRFENLKDPLKP"
BASE COUNT 450 a 362 c 494 g 462 t
ORIGIN
```

```
alignment_scores:
  Quality: 265.00      Length: 53
  Ratio: 5.096        Gaps: 0
Percent Similarity: 98.113 Percent Identity: 96.226

alignment_block:
US-09-209-961-21 x AF147751 ..
Align seg 1/1 to: AF147751 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIy 17
|||||
1016 ATGGTTTATTATTATTCATTTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1065

17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuValV 34
|||||
1066 ATTCCTGAATTGACATACATAGGTACACGCGATATTGTCTCGGTCG 1115

34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
|||||
1116 TATATCTGTTTTCGAACGCGCGTCCGAGGCGCTACGTGTCACATTCT 1165

51 ArgGlyLeu 53
|||||
1166 AGAGGTTTG 1174
```

OM of: US-09-209-961-21 to: N_Geneseq_36.* out_format : pfs
Date: Dec 28, 1999 1:50 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame1_p2n.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US09209961/runat.22121999_101043_17942/app_query.fasta.1
-DB=N_Geneseq_36 -Qfmt=fastap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-XGAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=biosum62
-TRANS=human40.cdi -LIST=45 -DOCLIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961
-NCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-209-961-21
Query length: 53
Database: N_Geneseq_36.*
Database sequences: 311585
Database length: 125096042
Search time (sec): 140.030000

score_list:	Sequence	Strid	Orig	ZScore	EScore	Len	Documentation
	N_Geneseq_36:X35213	+	273.00	689.13	8.8e-31	1768	! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35012	+	270.00	681.28	2.4e-30	1768	! Genomic DNA sequence of PCV st
	N_Geneseq_36:X35212	+	270.00	681.28	2.4e-30	1768	! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35214	+	270.00	681.28	2.4e-30	1768	! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35210	+	210.00	524.27	1.3e-21	1767	! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35211	+	210.00	524.27	1.3e-21	1767	! Nucleotide sequence of PCV iso
	N_Geneseq_36:X35013	-	68.00	152.71	0.6826	1759	! Nucleotide sequence which has
	N_Geneseq_36:Q34888	+	58.00	132.16	9.24	992	! Babesia equi probe 10. Nucleoti
	N_Geneseq_36:V32678	+	47.00	131.58	9.96	56	! Delta thioredoxin sequence PCR
	N_Geneseq_36:X41269	+	54.00	131.19	10.47	377	! Human secreted protein 5' EST
	N_Geneseq_36:Q29963	-	60.00	129.78	12.54	1106	! Sugar beet acidic chitinase SE
	N_Geneseq_36:Q29361	-	57.00	125.45	21.85	3350	! Ocean pout DNA sequence in opa
	N_Geneseq_36:V40339	+	56.50	124.53	24.29	1433	! Homo sapiens secreted protein
	N_Geneseq_36:T21462	+	50.00	124.52	24.64	256	! Human gene signature HUMGS02832
	N_Geneseq_36:V69410	+	55.50	123.87	26.76	1185	! Ehrlichia sp. HGE-2 3'-end DNA
	N_Geneseq_36:X11063	-	49.50	123.25	29.00	255	! Human biallelic polymorphic DNA
	N_Geneseq_36:T86283	-	53.50	122.78	30.78	777	! Chimeric Soybean albumin l/3 co
	N_Geneseq_36:V75227	+	51.00	122.75	30.89	400	! Staphylococcus aureus contig SE
	N_Geneseq_36:Q41226	+	59.00	122.35	32.54	3521	! Clone GP3 encoding major surfa
	N_Geneseq_36:X13828	+	50.50	121.37	36.87	403	! Enterococcus faecalis genome co
	N_Geneseq_36:X10557	+	46.00	121.08	38.27	125	! Human biallelic polymorphic DNA
	N_Geneseq_36:T13687	+	50.00	120.61	40.64	381	! ACNPV ORF 87, residues 74356-74
	N_Geneseq_36:V46398	+	52.00	120.31	42.26	670	! Human tumour suppressor TS10q23
	N_Geneseq_36:Q74199	+	45.00	119.90	44.53	108	! 3' end fragment of VJlms-BJ-NP
	N_Geneseq_36:Q40265	+	47.00	118.66	52.24	209	! Sequence of intermedate plasm
	N_Geneseq_36:V75559	+	49.50	118.61	52.55	409	! Staphylococcus aureus contig SE
	N_Geneseq_36:V69421	+	55.50	118.12	55.93	2129	! Ehrlichia sp. extended HGE-2 I
	N_Geneseq_36:X30735	+	52.00	117.96	57.10	851	! Streptococcus pneumoniae genom
	N_Geneseq_36:Q47735	+	50.50	117.77	58.56	582	! Hepatitis B core gene. Recombin
	N_Geneseq_36:V13060	-	45.50	117.73	58.85	154	! Interleukin 6 non-risk genotype
	N_Geneseq_36:Q60595	+	47.50	116.45	69.32	299	! Human brain Expressed Sequence
	N_Geneseq_36:V87840	+	47.00	116.25	71.10	267	! EST clone EY197. New polynucleo
	N_Geneseq_36:V32676	+	47.00	115.93	74.13	276	! Delta thioredoxin sequence. Pro
	N_Geneseq_36:V88392	+	48.00	115.62	77.16	372	! EST clone GP274. New polynucleo
	N_Geneseq_36:X20400	+	49.00	115.57	77.62	488	! Borrelia burgdorferi polynucleo
	N_Geneseq_36:T59139	-	42.00	115.24	80.93	78	! DNA polymerase ligand to a Therm
	N_Geneseq_36:V24930	-	49.00	115.19	81.44	507	! H. pylori cytoplasmic protein c
	N_Geneseq_36:Q77057	+	45.00	115.00	83.52	177	! Human genome fragment. New nucl
	N_Geneseq_36:T86282	+	50.50	114.93	84.25	777	! Soybean albumin 3 coding sequen
	N_Geneseq_36:V88484	+	46.00	114.40	90.17	247	! EST clone GN82. New polynucleo
	N_Geneseq_36:Q76403	+	47.00	114.08	93.89	333	! Human genome fragment. (prefer
	N_Geneseq_36:V89772	-	47.50	113.97	95.29	385	! EST clone CT207. New polynucleo

N_Geneseq_36:V22741	+	52.00	113.90	96.10	1287	! Babesia microti BMNI-11 ant
N_Geneseq_36:T91241	-	39.00	113.70	98.60	41	! Pseudomonas fluorescens lipas
N_Geneseq_36:X33156	-	48.00	113.43	102.17	465	! Potato isoamylase type DBE 3
seq_name: N_Geneseq_36:X35213						
seq_documentation_block:						
ID X35213 standard; DNA; 1768 BP.						
AC X35213;						
DT 01-JUL-1999 (first entry)						
DE Nucleotide sequence of PCV isolate Impl010.						
KW PCV isolate; type II porcine circovirus; PCV; PMWS;						
OS porcine multisystemic wasting syndrome; pig; vaccine; ss.						
QW Porcine circovirus.						
PN FR2769322-AL.						
PD 09-APR-1999.						
PF 22-JAN-1998; 000873.						
PR 03-OCT-1997; FR-012382.						
PA (MERI-) MERIAL SAS.						
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,						
FI Haines D, Harding J, Hassard L, Meehan B;						
DR WPI; 99-246948/21.						
PT New type II porcine circovirus, used for, e.g. passive immunization						
PS of pregnant sows						
Claim 14: Fig 4: 48pp; French.						
CC The present sequence represents the nucleotide sequence of PCV isolate						
CC Impl010. The specification describes a preparation of type II						
CC porcine circovirus (PCV), which is particularly isolated from a lesion,						
CC from a pig with symptoms of PMWS (porcine multisystemic wasting						
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from						
CC it, and vectors that express these polypeptides are all useful in						
CC vaccines, suitable for administration to adult or young pigs, or to						
CC pregnant sows (for passive immunization of their offspring). DNA						
CC isolated from PCV is used for in vivo or in vitro expression of viral						
CC polypeptides, also as probes or primers for diagnosis in usual						
CC hybridization or amplification assays. These polypeptides may also be						
CC used diagnostically to detect PCV-specific antibodies, while antibodies						
CC raised against the polypeptides can be used to detect antigens, in any						
CC usual immunoassay format.						
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;						
alignment_scores:						
Quality: 273.00 Length: 53						
Ratio: 5.151 Gaps: 0						
Percent Similarity: 100.000 Percent Identity: 100.000						
alignment_block:						
US-09-209-961-21 x X35213						
Align seg 1/1 to: X35213 from: 1 to: 1768						
1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLeu 17						
1363 ATGGTTTATTATTATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1412						
17 sPheSerGluLeuTyIleHisGlyTyThrAspIleValValLeuVal 34						
1413 ATTCCTGAATTGATACATACATGTTTACACGGATATTAGTCTGTGTCG 1462						
34 alphaThrValPheGluArgSerAlaGluAlaTyrrValValHisIleSer 50						
1463 TATTACTGTTTTCGAACGACGCGCGAGCGCTACGTGGTCCACATTTC 1512						
51 ArgGlyLeu 53						
1513 AGAGGTTTG 1521						
seq_name: N_Geneseq_36:X35012						
seq_documentation_block:						
ID X35012 standard; DNA; 1768 BP.						
AC X35012;						

DT 01-JUL-1999 (first entry)
 DE Genomic DNA sequence of PCV strain 999PCV.
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
 KW vaccine; ss.
 OS Porcine circovirus
 PD FR2769321-Al.
 PN 09-APR-1999.
 PF 03-OCT-1997; 012382.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246948/21.
 PT New porcine circovirus from animals with porcine systemic wasting
 PT syndrome
 PS Claim 13; Fig 1; 35pp; French.
 CC The present sequence represents the genomic sequence of porcine
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

 alignment_scores:
 Quality: 270.00 Length: 53
 Ratio: 5.094 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.113

 alignment_block:
 US-09-209-961-21 x X35012 ..
 Align seg 1/1 to: X35012 from: 1 to: 1768

 1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
 |||||
 1364 ATGGTTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1413

 17 sPheSerGlulLeuTyrlleHisGlyTyrlleHisGlyValPheLysIleSer 34
 |||||
 1414 ATTCCTCGAATTGTACATACATGTTACACGGATATTGTAGTCTGGTCG 1463

 34 alPheThrValPheGluArgSerAlaGluAlaTyrlleHisIleSer 50
 |||||
 1464 TATATCTGTTTTCGAACGACGAGTCCGAGGCTACGTCGTCACATTTC 1513

 51 ArgGlyLeu 53
 |||||
 1514 AGAGGTTTG 1522

 seq_name: N_Geneseq_36:X35212

 seq_documentation_block:
 ID X35212 standard; DNA; 1768 BP.
 AC X35212;
 DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
 DE PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 03-OCT-1997; 000873.
 PR (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows

PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PA Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows
 PS Claim 14; Fig 3; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Imp999. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

 alignment_scores:
 Quality: 270.00 Length: 53
 Ratio: 5.094 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.113

 alignment_block:
 US-09-209-961-21 x X35212 ..
 Align seg 1/1 to: X35212 from: 1 to: 1768

 1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
 |||||
 1363 ATGGTTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1412

 17 sPheSerGlulLeuTyrlleHisGlyTyrlleHisGlyValPheLysIleSer 34
 |||||
 1413 ATTCCTCGAATTGTACATACATGTTACACGGATATTGTAGTCTGGTCG 1462

 34 alPheThrValPheGluArgSerAlaGluAlaTyrlleHisIleSer 50
 |||||
 1463 TATATCTGTTTTCGAACGACGAGTCCGAGGCTACGTCGTCACATTTC 1512

 51 ArgGlyLeu 53
 |||||
 1513 AGAGGTTTG 1521

 seq_name: N_Geneseq_36:X35214

 seq_documentation_block:
 ID X35214 standard; DNA; 1768 BP.
 AC X35214;
 DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
 DE PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows

PS Claim 14; Fig 6; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Imp99. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 CC Sequence 1768 Bp; 453 A; 359 C; 486 G; 455 T;

alignment_scores:
 Quality: 270.00 Length: 53
 Ratio: 5.094 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 98.113
 alignment_block:
 US-09-209-961-21 x X35214 ..
 Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
 |||||
 1364 ATGGTTTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1413
 17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValV 34
 |||||
 1414 ATTCCTGAATTGTACATACATGTTACACGGATATTGATCTCGTGGTCG 1463
 34 alPheThrValPheGluArgSerAlaGluAlaTyrlValValHisIleSer 50
 |||||
 1464 TATATACTGTTTTCGAACGACGTGCGGAGCGCTACGTGTCACATTTCT 1513

51 ArgGlyLeu 53
 |||||
 1514 AGAGGTTTG 1522

seq_name: N_Geneseq_36:X35210

seq_documentation_block:
 ID X35210 standard; DNA; 1767 BP.
 AC X35210;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl011-48121.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERTAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows
 PS Claim 14; Fig 1; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl011-48121. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA

CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 CC Sequence 1767 Bp; 447 A; 360 C; 502 G; 458 T;

alignment_scores:
 Quality: 210.00 Length: 53
 Ratio: 4.200 Gaps: 0
 Percent Similarity: 94.340 Percent Identity: 86.792
 alignment_block:
 US-09-209-961-21 x X35210 ..
 Align seg 1/1 to: X35210 from: 1 to: 1767

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLy 17
 |||||
 1363 ATGGTTTTTATTATTCATTAGGTTTAAAGTGGGGGCTTTAAGATTAA 1411
 17 sPheSerGluLeuTyrlleHisGlyTyThrAspIleValValLeuValV 34
 |||||
 1412 ATTCCTGAATTGTACATACATGTTACACGGATATTGATCTCGTGGTCG 1461
 34 alPheThrValPheGluArgSerAlaGluAlaTyrlValValHisIleSer 50
 |||||
 1462 TATATACTGTTTTCGAACGACGTGCGGAGCGCTACGTGTCACATTTCC 1511

51 ArgGlyLeu 53
 |||||
 1512 AGCAGTTTG 1520

seq_name: N_Geneseq_36:X35211

seq_documentation_block:
 ID X35211 standard; DNA; 1767 BP.
 AC X35211;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl011-48285.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-Al.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERTAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows
 PS Claim 14; Fig 2; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl011-48285. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 CC Sequence 1767 Bp; 448 A; 359 C; 500 G; 460 T;

alignment_scores:
 Quality: 210.00 Length: 53
 Ratio: 4.200 Gaps: 0
 Percent Similarity: 94.340 Percent Identity: 86.792

alignment_block:
 US-09-209-961-21 x X35211 ..

Align seg 1/1 to: X35211 from: 1 to: 1767

1 MetValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleIy 17
 |||||
 1363 ATGCTTTTATTATTAAGGTT.AAGTGGGGGCTTTAAGATTAA 1411
 |||||
 17 sPheSerGluLeuIleHisGlyTyrThrAspIleValValLeuValV 34
 |||||
 1412 ATTCTGAATGTACATACATAGTTACACGATATTATTCCTGGTGG 1461
 |||||
 34 alPheThrValPheGluArgSerAlaGluAlaTyrValValHisIleSer 50
 |||||
 1462 TATATCTGTTTTCGAGCGAGTCCGAGGCTAGTGTCTACATTCC 1511
 |||||
 51 ArgGlyLeu 53
 |||||
 1512 AGTAGTTTG 1520

seq_name: N_Geneseq_36.X35013

seq_documentation_block:
 ID X35013 standard; DNA; 1759 BP.
 AC X35013;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence which has homology to PCV sequence.
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
 KW vaccine; ss.
 OS Sus sp.
 PN FR2769321-AL.
 PD 09-APR-1999.
 PF 03-OCT-1997; 012382.
 PR 03-OCT-1997; FR-012382.
 PA (NERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246947/21.

PT New porcine circovirus from animals with porcine systemic wasting
 PT syndrome
 PS Disclosure; Fig 2; 35pp; French.
 CC The specification describes a genomic sequence of porcine
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format. The present sequence appears in the
 CC specification.
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment_scores:
 Quality: 68.00 Length: 41
 Ratio: 2.519 Gaps: 0
 Percent Similarity: 65.854 Percent Identity: 43.902

alignment_block:
 US-09-209-961-21 x X35013/rev ..

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

13 ValPheLysIleLysPheSerGluLeuTyrIleHisGlyTyrThrAspI1 29
 |||||
 375 GTCTTTTAGGATAAATCTCTGAATTGTACATAAATAGTCAGCCTTACCA 326
 |||||
 29 eValValLeuValValPheThrValPheGluArgSerAlaGluAlaTyrV 46
 |||||
 325 CATAAATTTTGGCTGTGGCTGCAATTTGGAGCGCATAGCCGAGGCTGTG 276
 |||||
 46 alValHisIleSerArgGlyLeu 53
 |||||
 275 TCTCGACATTTGGTGTGGTATT 253

seq_name: N_Geneseq_36:Q34688

seq_documentation_block:
 ID Q34688 standard; DNA; 992 BP.
 AC Q34688; 1993 (first entry)
 DT 10-MAY-1993
 DE Babesia equi probe 10.
 KW Probe; screen; infected animal; carrier; tick; vector; cDNA library;
 KW plasmid; pUC13; horse; high specificity; protozoan; repetitive DNA; ss.
 OS Babesia equi.
 PN ZA910544S-A.
 PD 29-APR-1992.
 PF 12-JUL-1991; 005449.
 PR 20-APR-1990; ZA-003002.
 PA (TECH-) TECHNOLOGY FINANCE CORP PTY LTD.
 PI Ambrosio RE, Posnett ES;
 DR WPI; 93-018413/02.
 PT Nucleotide probes for detecting Babesia equi in blood - by
 PT hybridising with portion of genome of Babesia equi etc.
 PS Claim 17; Fig 7; 24pp; English.
 CC This sequence is a portion of the B. equi genome which was used to
 CC as a probe to detect the presence of B. equi. Further probe sequences
 CC of the invention are given in Q34678-86. These probes can be used to
 CC screen animals for the presence of Babesia equi and to identify
 CC infected animals and carriers of the disease. They may also be used
 CC to detect Babesia parasites in the tick vectors. The probes were
 CC isolated, using standard techniques from a cDNA library of B. equi
 CC which was constructed in the plasmid vector, pUC13. The library was
 CC screened and the clones which did not cross-hybridise to horse DNA were
 CC investigated further. The resulting probes show high specificity for
 CC the particular protozoan species and can be synthesised in multiple
 CC copies. These probes are based on a repetitive DNA sequence which
 CC occurs a number of times in the genome of B. equi.
 SQ Sequence 992 BP; 259 A; 195 C; 197 G; 339 T;

alignment_scores:
 Quality: 58.00 Length: 55
 Ratio: 1.871 Gaps: 2
 Percent Similarity: 56.364 Percent Identity: 30.909

alignment_block:
 US-09-209-961-21 x Q34688 ..

Align seg 1/1 to: Q34688 from: 1 to: 992

3 PheIleIleHisLeuGlyPheLysTrpGly.....ValPh 14
 |||||
 485 TTGTCACTACATTAGGATTACACTTGTGGCTTTGCCCATTTATGTT 534
 |||||
 14 elysIleLysPheSerGluLeuTyrIleHisGly..... 25
 |||||
 535 TTCTGTGAGTTTTTACCATGACCTTTTACATGCGCGGTAGGAAGGGGTGG 584
 |||||
 26TyrThrAspIleValValLeuValValPheThrValPheGluArg 40
 |||||


```

FT WO9217591-A. /*tag= d
PN
PD 15-OCT-1992.
PF 07-APR-1992; DK0108.
PR 08-APR-1991; DK-000616.
PA (DANI-) DANISCO AS.
PI Berglund L, Bojsen K, Mikkelsen JD, Nielsen KK;
DR WPI: 92-366261/44.
DR P-PSDB; R28148.
PT Sugar beet chitinase 4 and corresponding DNA - inhibits
PT germination and growth of chitin-containing fungi and is used to
PT produce genetically transformed plants
PS Claim 26; Page 158-159; 254pp; English.
CC A clone encoding the acidic chitinase SE gene was isolated from a
CC sugar beet lambda ZAP cDNA library. The sequence encodes a protein
CC having 293 amino acid residues. The leader sequence consists of 25
CC amino acids and the functional domain of 268 amino acid residues.
CC The leader sequence of acidic chitinase SE is a preferred leader
CC sequence for the constructs of the invention. The constructs encode
CC chitinases and can be used in the production of transgenic plants
CC having increased resistance to nematodes and chitin-containing
CC plant pathogens, partic. fungi.
SQ Sequence 1106 BP; 313 A; 251 C; 226 G; 316 T;

alignment_scores:
Quality: 57.50 Length: 35
Ratio: 2.212 Gaps: 1
Percent Similarity: 74.286 Percent Identity: 31.429

alignment_block:
US-09-209-961-21 x Q29963 ..
Align seg 1/1 to: Q29963 from: 1 to: 1106
1 MetValPheLeuIleHisLeuGlyPheLysTrpGlyValPheLysIleLY 17
:::|||||:|||||:|||||:|||||:|||||:
999 CTGTGTAATACATACTATTATATCATAAATGGCTTATGCTGCGTATA 1048
|||||:|||||:|||||:|||||:|||||:
17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValValV 34
|||||:|||||:|||||:|||||:|||||:
1049 TTTTGGTGATTATATATACACAGTTAC...AACTGCGAATTATGCGAG 1095
34 alPhe 35
|||||
1096 TCTTT 1100

seq_name: N_Geneseq_36:Q29361
seq_documentation_block:
ID Q29361 standard; cDNA; 3350 BP.
AC Q29361;
DT 08-MAR-1993 (first entry)
DE Ocean pout DNA sequence in opAPP-V.
KW GH; transgenic; fish; antifreeze protein promoter; growth hormone; ss.
OS Oncorhynchus keta.
FH Key Location/Qualifiers
FT misc_feature 1004..1020
FT /*tag= a
FT /*note= "liver-specific sequence"
FT misc_feature 1944..1960
FT /*tag= b
FT /*note= "liver-specific sequence"
FT caat_signal 2006..2010
FT /*tag= c
FT tata_signal 2049..2053
FT /*tag= d
FT terminator 2253..2258
FT /*tag= e
FT terminator 2285..2291
FT /*tag= f
PN WO9216618-A.
PD 01-OCT-1992.

```

```

PF 12-MAR-1992; CA0109.
PR 15-MAR-1991; US-669765.
PA (HSCR-) HSC RES & DEV LP.
PA (SEAB-) SEABRIGHT CORP LTD.
PI Fletcher GL, Hew CL;
DR WPI: 92-349205/42.
PT Gene construct for prodn. of transgenic fish - contains
PT antifreeze protein promoter
PS Claim 2; Page 49; 70pp; English.
CC The cDNA sequence for the antifreeze protein of Ocean pout may be
CC used in a chimeric gene construct incorporation into fish genome
CC to produce a transgenic fish with increased expression of a gene of
CC interest e.g. the growth hormone gene, thus causing a dramatic increase
CC in weight and growth rate. The construct is expressed predominantly in
CC the liver and its expression is not affected by the host genome.
CC See also Q29359-65.
SQ Sequence 3350 BP; 1014 A; 633 C; 650 G; 1052 T;

alignment_scores:
Quality: 60.00 Length: 36
Ratio: 2.500 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 38.889

alignment_block:
US-09-209-961-21 x Q29361/rev ..
Align seg 1/1 to reverse of: Q29361 from: 1 to: 3350
23 IleHisGlyTyrThrAspIleValValLeuValValPheThrValPheG1 39
|||||:|||||:|||||:|||||:|||||:
2981 ATTCAATAATTACATTGTTACTTCTCTCCGATCTTACTTCTTATAA 2932

39 u.ArgSerAlaGluAla.....TyrValVal 47
: ||||| |||||
2931 CACGTCCTCATGAGCGCATCTCAATAAATAAATGTATATATATATA 2882
48 HisIle 49
|||||
2881 CATATA 2876

seq_name: N_Geneseq_36:V40539
seq_documentation_block:
ID V40539 standard; cDNA; 1433 BP.
AC V40539;
DT 09-NOV-1998 (first entry)
DE Homo sapiens secreted protein clone AA35_2.
KW Clone; secreted protein; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 506..646
FT /*tag= a
FT /*note= "secreted protein"
PN
PD 30-JUL-1998.
PR 24-JAN-1997; U01396.
PR 24-JAN-1997; US-788789.
PA (GEM) GENETICS INST INC.
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Racie LA, Spaulding V, Treacy M;
DR WPI: 98-427949/36.
DR P-PSDB; W29669.
PT New isolated polynucleotide(s) and secreted proteins - isolated from
PT human foetal kidney, adult brain, adult salivary gland, foetal brain
PT and adult testes cDNA libraries
PS Claim 13; Page 63; 109pp; English.
CC The sequence is that of encoding a secreted protein. Such a
CC protein can have biological activities, e.g. nutritional
CC activity, cytokine and cell proliferation/differentiation activity,
CC immune stimulating or suppressing activity, haematopoiesis regulating
CC activity, tissue growth activity, activin/inhibin activity,
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,

```

CC	recognising different cell types.
SQ	Sequence 256 BP; 86 A; 26 C; 42 G; 99 T;

6 HisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerGlu..... 20
||||| ::|||::| ||| ::||| ::|
359 CATCTCCACTGGAAATGG.....TTTCACCTACCAATTCTGTAATCACTT 316
21LeuTyrlle...HisGlyTyrThrAspIleValLeuValValp 35
||||| ::| ||| |||||::|||::| ::|
315 AGCTCTATATCTATACCATACATATACGCAAAATCCTCTTTAATCCCTCT 266
35 heThrValPheGluArgSerAlaGluLayrValValHisIleSerArg 51
::| ||| ::|||::| ::| ::| ::| ::| ::| ::| ::|
265 AC..GTTCACGACATTTGTTTAAAGTGCACAGATACCACTTAAGAGA 219


```
seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-618-911-3
seq_documentation_block:
; Sequence 3, Application US/08618911
; Patent No. 5850016
; GENERAL INFORMATION:
; APPLICANT: Jung, Rudolf
; APPLICANT: Hastings, Craig
; APPLICANT: Coughlan, Sean
; APPLICANT: Hu, David
; TITLE OF INVENTION: ALTERATION OF AMINO ACID COMPOSITIONS IN
; TITLE OF INVENTION: SEEDS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PIONEER HI-BRED INTERNATIONAL, INC.
; STREET: 700 Capital Square, 400 Locust Street
; CITY: Des Moines
; STATE: Iowa
; COUNTRY: USA
; ZIP: 50309
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,911
; FILING DATE: Concurrently herewith
; ATTORNEY/AGENT INFORMATION:
; NAME: Simon, Soma
; REGISTRATION NUMBER: 37,444
; REFERENCE/DOCKET NUMBER: 365-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (515) 248-4800
; TELEFAX: (515) 248-4844
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 777 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 28..501
; US-08-618-911-3

alignment_scores:
Quality: 50.50 Length: 39
Ratio: 2.020 Gaps: 1
Percent Similarity: 64.103 Percent Identity: 25.641
alignment_block:
US-09-209-961-21 x US-08-618-911-3/rev ..
Align seg 1/1 to reverse of: US-08-618-911-3 from: 1 to: 777
4 IleileHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerG1 20
:::|||||:::|||||:::
315 CRTCTGCATGTGTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 266
20 uLeuTyrlle.....HisGlyTyThrAspIleValVall 32
|||||
265 TCCTTATGTAGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTG 216
32 euValValPheThrVal 37
215 TCGTCTCATCATCATC 199
seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US96-09451-37
seq_documentation_block:
; Sequence 3, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-464-164-3
seq_documentation_block:
; Sequence 3, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-464-164-3
seq_documentation_block:
; Sequence 37, Application PC/TUS9609451
; GENERAL INFORMATION:
; APPLICANT: LARRY GOLD
; APPLICANT: SUMEDHA JAYASENA
; TITLE OF INVENTION: NUCLEIC ACID LIGAND INHIBITORS TO
; TITLE OF INVENTION: DNA POLYMERASES
; NUMBER OF SEQUENCES: 77
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Swanson and Bratschun, L.L.C.
; STREET: 8400 East Prentice Ave., Suite 200
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09451
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,426
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/487,720
; FILING DATE: 7-JUNE-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/484,557
; FILING DATE: 7-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Barry J. Swanson
; REGISTRATION NUMBER: 33,215
; REFERENCE/DOCKET NUMBER: NEX43C/PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 793-3333
; TELEFAX: (303) 793-3433
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 78 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; PCT-US96-09451-37

alignment_scores:
Quality: 42.00 Length: 14
Ratio: 3.500 Gaps: 0
Percent Similarity: 85.714 Percent Identity: 57.143
alignment_block:
US-09-209-961-21 x PCT-US96-09451-37/rev ..
Align seg 1/1 to reverse of: PCT-US96-09451-37 from: 1 to: 78
19 SerGluLeuTyrlleHisGlyTyThrAspIleValVall 32
|||||
72 AGCGAATCATACACAGAGGATCTCTGACATCACTACTG 31
seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-464-164-3
seq_documentation_block:
; Sequence 3, Application US/08464164
; Patent No. 5614195
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
```



```

: APPLICANT: Bumstead, Janene M.
: APPLICANT: Vermeulen, Arno N.
: TITLE OF INVENTION: Coccidiosis poultry vaccine
: NUMBER OF SEQUENCES: 10
: CORRESPONDENCE ADDRESS:
: ADDRESS: Akzo No. 5614195el Patent Department
: STREET: 1300 Piccard Drive, Suite 206
: CITY: Rockville
: STATE: Maryland
: COUNTRY: U.S.A.
: ZIP: 20850
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/464,164
: FILING DATE: June 2, 1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Gormley, Mary E.
: REGISTRATION NUMBER: 34,409
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (301) 258-5200
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 242 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cdna to mRNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: FRAGMENT TYPE: C-terminal
: ORIGINAL SOURCE:
: ORGANISM: Eimeria maxima
: STRAIN: Houghton
: DEVELOPMENTAL STAGE: Sporozoite
: IMMEDIATE SOURCE:
: LIBRARY: Sporozoite cdna cloned in Lambda ZAPII
: CLONE: Em70-4, 5' end of clone
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 3..242
: US-08-464-164-3

alignment_scores:
    Quality: 45.00      Length: 26
    Ratio: 2.500        Gaps: 1
Percent Similarity: 69.231 Percent Identity: 42.308

alignment_block:
US-09-209-961-21 x US-08-464-164-3  ..

Align seg 1/1 to: US-08-464-164-3 from: 1 to: 242

25 GlyThrAspIleValValLeuValVal.PheThrValPheGluAr9S 41
||| |||||
117 GTCACCTTTGGTATGATCCGTCCTGTGTTATGACAAACGGTGTATTCGAAGT 166
41 erAlaGluAlaTyValValHisrle 49
|| :|||
167 CCATGCACCTTCAGTGATACATC 192

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-338-057-3

seq_documentation_block:
: Sequence 3, Application US/08338057
: Patent No. 5795741
: GENERAL INFORMATION:
: APPLICANT: Tomley, Fiona M.

```

```
; Sequence 3, Application US/08668416
; Patent No. 5843722
; GENERAL INFORMATION:
; APPLICANT: Tomley, Fiona M.
; APPLICANT: Dunn, Paul P. J.
; APPLICANT: Bumstead, Janene M.
; APPLICANT: Vermeulen, Arno N.
; TITLE OF INVENTION: Coccidiosis poultry vaccine
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Akzo No. 5843722el Patent Department
; STREET: 1300 Piccard Drive, Suite 206
; CITY: Rockville
; STATE: Maryland
; COUNTRY: U.S.A.
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,416
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,164
; FILING DATE: June 2, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gormley, Mary E.
; REGISTRATION NUMBER: 34,409
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 258-5200
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 242 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Eimeria maxima
; STRAIN: Houghton
; DEVELOPMENTAL STAGE: Sporozoite
; IMMEDIATE SOURCE:
; LIBRARY: sporozoite cDNA cloned in Lambda ZAPII
; CLONE: Em70-4, 5'end of clone
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..242
; US-08-668-416-3

alignment_scores:
  Quality: 45.00      Length: 26
  Ratio: 2.500        Gaps: 1
  Percent Similarity: 69.231  Percent Identity: 42.308

alignment_block:
  US-09-209-961-21 x US-08-668-416-3
  Align seg 1/1 to: US-08-668-416-3 from: 1 to: 242

      25 GlyThrAspIleValValLeuValVal.PheThrValPheGluArgS 41
      117 GGTACCTTTGATGTATCGCTCTTTATTGACACCGGTGATTGCAAGT 166
      41 erAlaGluAlaTyrValValHisIle 49
      167 QCATGCAACTTCAGGTGATACATC 192
```

```
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-477-559-17
seq_documentation_block:
; Sequence 17, Application US/08477559
; Patent No. 5723765
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trolinder, No. 5723765ma L. G.
; APPLICANT: Kelm, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, P.C.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,559
; FILING DATE: 01-AUG-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-559-17

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000        Gaps: 0
  Percent Similarity: 66.667  Percent Identity: 47.619

alignment_block:
  US-09-209-961-21 x US-08-477-559-17/rev
  Align seg 1/1 to reverse of: US-08-477-559-17 from: 1 to: 113

      14 PheLysIleLysPheSerGluLeuTyrIleHisGlyTyrThrAspIleVa 30
      84 TTCACCTCTATCAGTCATAGATGCTTTATATACACTCTATCATGATAGACT 35
      30 lValLeuValVal 34
      34 GGGATTGTGCGTC 22

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-995-161-17
seq_documentation_block:
; Sequence 17, Application US/08995161
; Patent No. 5925808
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
```

```

; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trolinder, No. 5925808ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, p.c.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,161
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/477,559
; FILING DATE: 01-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 113 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-995-161-17

```

```

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 47.619

alignment_block:
US-09-209-961-21 x US-08-995-161-17/rev ..

Align seg 1/1 to reverse of: US-08-995-161-17 from: 1 to: 113

14 PheylsileLysPheSerGluLeuTyrlleHisGlyTyThrAspileVa 30
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
84 TTCACCTATCATCATAGATAGATCTTATATACACTATCATCATAGACT 35

30 lValLeuValVal 34
| ||| |||
34 GGGATTGGCGTC 22

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-475-891A-3

seq_documentation_block:
; Sequence 3, Application US/08475891A
; Patent No. 5859339
; GENERAL INFORMATION:
; APPLICANT: Ronald, Pamela C.
; APPLICANT: Wang, Guo-Liang
; APPLICANT: Song, Wen-Yuang
; TITLE OF INVENTION: Procedures and Materials for Conferring
; TITLE OF INVENTION: Disease Resistance in Plants
; NUMBER OF SEQUENCES: 15

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/475,891A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/373,375
; FILING DATE: 17-JAN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bastian, Kevin L.
; REGISTRATION NUMBER: 34,774
; REFERENCE/DOCKET NUMBER: 02370-0589100S
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5992 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(512..3149, 3993..4393)
; OTHER INFORMATION: /product= "RRK-B"
; OTHER INFORMATION: /note= "Xa21 Xanthomonas spp. disease
; OTHER INFORMATION: resistance gene RRK-B from rice (Oryza
; OTHER INFORMATION: sativa)"
; US-08-475-891A-3

```

```

alignment_scores:
  Quality: 56.50      Length: 55
  Ratio: 2.093       Gaps: 1
  Percent Similarity: 49.091   Percent Identity: 32.727

alignment_block:
US-09-209-961-21 x US-08-475-891A-3 ..

Align seg 1/1 to: US-08-475-891A-3 from: 1 to: 5992

4 llelleHisleuGlyPheLysTrpGlyValPheLysleLysPheSerGl 20
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5313 ATATGTTTCTATAGTTTAAATGTTCTGTATTTGTACCGTTTCNGA 5362

20 uLeuTyrlleHis.....Glyt 26
:||||| |||
5363 TCTGTACCGACATATTTCCATCAGTATTATTCATTTCCGTTTCCGATA 5412

26 yThrAspileValValLeuValPheThrValPheGluArgSerAla 42
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
5413 TTTCGATATCGTTTTCGTTTCGACTTTACCGTTTCGATTTTCATTCC 5462

43 GluAlaTyrrValVal 47
||| :|||
5463 GAGAAATATGATT 5477

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-477-559-18
seq_documentation_block:
; Sequence 18, Application US/08477559

```

```
; Patent No. 5723765
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trollinger, No. 5723765ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, p.c.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,559
; FILING DATE: 01-AUG-1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-477-559-18

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 47.619

US-09-209-961-21 x US-08-477-559-18 ..
Align seg 1/1 to: US-08-477-559-18 from: 1 to: 117

alignment_block:
US-09-209-961-21
14 PhelysileLysPheSerGluLeuTyrIleHisGlyTyrThrAspIleVa 30
   ||| : : : : : : : ||| ||| ||| ||| : : : ||| |||
34 TTCACCTATCATCAGTAGAGCTCTTATATACACTCTATCATCTATCATGATAGT 83

seq_name: /cqn2_6/ptodata/1/ina/5D_COMB.seq:US-08-995-161-18

seq_documentation_block:
; Sequence 18, Application US/08995161
; Patent No. 5925808
; GENERAL INFORMATION:
; APPLICANT: Oliver, Melvin J.
; APPLICANT: Quisenberry, Jerry E.
; APPLICANT: Trollinger, No. 5925808ma L. G.
; APPLICANT: Keim, Don L.
; TITLE OF INVENTION: Control of Plant Gene Expression
; NUMBER OF SEQUENCES: 26
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rothwell, Figg, Ernst & Kurz, p.c.
; STREET: 555 Thirteenth Street, N. W.
; CITY: Washington
; STATE: D. C.
; COUNTRY: USA
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/995,161
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 08/477,559
; APPLICATION NUMBER: 01-AUG-1994
; FILING DATE: 01-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Figg, E. Anthony
; REGISTRATION NUMBER: 27,195
; REFERENCE/DOCKET NUMBER: 1760-174A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 783-6040
; TELEFAX: (202) 783-6031
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 117 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; US-08-995-161-18

alignment_scores:
  Quality: 42.00      Length: 21
  Ratio: 3.000       Gaps: 0
  Percent Similarity: 66.667   Percent Identity: 47.619

US-09-209-961-21 x US-08-995-161-18 ..
Align seg 1/1 to: US-08-995-161-18 from: 1 to: 117

alignment_block:
US-09-209-961-21
14 PhelysileLysPheSerGluLeuTyrIleHisGlyTyrThrAspIleVa 30
   ||| : : : : : : : ||| ||| ||| ||| : : : ||| |||
34 TTCACCTATCATCAGTAGAGCTCTTATATACACTCTATCATCTATCATGATAGT 83

seq_name: /cqn2_6/ptodata/1/ina/5B_COMB.seq:US-08-122-433-9

seq_documentation_block:
; Sequence 9, Application US/08122433
; Patent No. 5683985
; GENERAL INFORMATION:
; APPLICANT: Chu, Barbara C.F.
; APPLICANT: Orgel, Leslie
; TITLE OF INVENTION: OLIGODEOXYNUCLEOTIDES AND
; TITLE OF INVENTION: OLIGONUCLEOTIDES USEFUL AS DECOYS FOR PROTEINS WHICH
; TITLE OF INVENTION: SELECTIVELY BIND TO DEFINED DNA SEQUENCES
; NUMBER OF SEQUENCES: 47
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
```

```
;
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,433
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/687,337
; FILING DATE: 18-APR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Reiter, Stephen E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P31 9308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-546-9392
; TELEFAX: 619-546-9392
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 55 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; US-08-122-433-9

alignment_scores:
  Quality: 39.00      Length: 9
  Ratio: 4.875       Gaps: 0
  Percent Similarity: 88.889   Percent Identity: 66.667

alignment_block:
  US-09-209-961-21 x US-08-122-433-9/rev ..
  Align seg 1/1 to reverse of: US-08-122-433-9 from: 1 to: 55

  9 PhelystirpGlyValPheLysIleLys 17
  |||::|||::|||::|||::|||::|||
  48 TTCGCTGGGACTTTTCARAAAGAA 22

seq_name: /cgn2.6/ptodata/1/ina/5A_COMB.seq:US-08-308-892A-16

seq_documentation_block:
; Sequence 16, Application US/08308892A
; Patent No. 5500341
; GENERAL INFORMATION:
; APPLICANT: Spears, Patricia A.
; TITLE OF INVENTION: SPECIES-SPECIFIC DETECTION OF
; TITLE OF INVENTION: MYCOBACTERIUM KANSASII
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Richard J. Rodrick, Becton Dickinson and
; ADDRESSEE: Company
; STREET: 1 Becton Drive
; CITY: Franklin Lakes
; STATE: NJ
; COUNTRY: US
; ZIP: 07417
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,892A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
```

```
;
; NAME: Fugit, Donna R.
; REGISTRATION NUMBER: 32,135
; REFERENCE/DOCKET NUMBER: P-3128
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 171 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-308-892A-16

alignment_scores:
  Quality: 43.00      Length: 27
  Ratio: 2.263       Gaps: 0
  Percent Similarity: 70.370   Percent Identity: 29.630

alignment_block:
  US-09-209-961-21 x US-08-308-892A-16/rev ..
  Align seg 1/1 to reverse of: US-08-308-892A-16 from: 1 to: 171

  11 TrpGlyValPheLysIleLysPheSerGluLeuTyrIleHisGlyTyrTh 27
  |||::|||::|||::|||::|||::|||
  113 TGGGCGCCGAGGCGCTCGAGTTCGCCGAGGTGCTGTCGACGATATCC 64

  27 RasPleValValLeuValPheThrVal 37
  ::::::::::::::::::::|||::|
  63 CAGCGTCGCGCTGTTGCCGCCGCTACTGTC 33

seq_name: /cgn2.6/ptodata/1/ina/5C_COMB.seq:US-08-481-337A-3

seq_documentation_block:
; Sequence 3, Application US/08481337A
; Patent No. 5863738
; GENERAL INFORMATION:
; APPLICANT: TEN DIJKE, Peter
; APPLICANT: HELDIN, Carl-Henrik
; APPLICANT: MIYAZONO, Komei
; APPLICANT: SAMPATH, Kuber T.
; TITLE OF INVENTION: Morphogenic Protein-Specific Cell
; TITLE OF INVENTION: Surface Receptors and Uses Therefor
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Testa, Hurwitz & Thibault
; STREET: 125 High St.
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/481,337A
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: MEYERS, Thomas C.
; REGISTRATION NUMBER: 36,989
; REFERENCE/DOCKET NUMBER: CRP-097CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1630
; OTHER INFORMATION: /product= "Human ALK2"
US-08-481-337A-3

alignment_scores:
  Quality: 53.00      Length: 27
  Ratio: 2.524        Gaps: 0
  Percent Similarity: 77.778      Percent Identity: 40.741

alignment_block:
US-09-209-961-21 x US-08-481-337A-3 ..
Align seg 1/1 to: US-08-481-337A-3 from: 1 to: 2724
20 GluLeuTyrIleHisGlyTyrThrAspIleValValLeuValValPheTh 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2250 GAATTGTTTATACACAACCTTTGCAAAATTATTATTACTTGTGCACCTAGT 2299
36 rValPheGluArgSerAlaGluAlaTyrVal 46
:|||||: :|||:|||||:
2300 AGTTTTCACAAAACCTGCTTTGTGCATATGTT 2330

seq_name: /cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-05467-3

seq_documentation_block:
; Sequence 3, Application PC/TUS9505467
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: MORPHOGENIC PROTEIN-SPECIFIC CELL
; TITLE OF INVENTION: SURFACE RECEPTORS AND USES THEREFOR
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
; ADDRESSEE: THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05467
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER, EDMUND R.
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: CRP-097PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2724 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 104..1633
; OTHER INFORMATION: /product= "Human ALK2"
PCT-US95-05467-3

alignment_scores:
  Quality: 53.00      Length: 27
  Ratio: 2.524        Gaps: 0
  Percent Similarity: 77.778      Percent Identity: 40.741

alignment_block:
US-09-209-961-21 x PCT-US95-05467-3 ..
Align seg 1/1 to: PCT-US95-05467-3 from: 1 to: 2724
20 GluLeuTyrIleHisGlyTyrThrAspIleValValLeuValValPheTh 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2250 GAATTGTTTATACACAACCTTTGCAAAATTATTATTACTTGTGCACCTAGT 2299
36 rValPheGluArgSerAlaGluAlaTyrVal 46
:|||||: :|||:|||||:
2300 AGTTTTCACAAAACCTGCTTTGTGCATATGTT 2330
```



```

: ||| ... :|||...:|||||
122 GAAGGGACTAGT...GTGGTCGGTGTAGTGTCTCAGGGAAGTGTCAACT 76
40 rgSerAlaGluAlaTyrValValHisIleSerArg 51
|||||...:|||||...:|||||
75 CATCCGCAGAGAATTCCTGCAGCATCTCCAGAAA 41

```

seq_name: gb_gss9:A0620563

```

seq_documentation_block:
LOCUS A0620563 575 bp DNA GSS 16-JUN-1999
DEFINITION HS-5190_B1_C02_T7A RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate=766 Col=3 Row=F, genomic survey sequence.
ACCESSION A0620563
NID 95082955
VERSION A0620563.1 GI:5082955
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 575)
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
Shaker,R., Schmidt,S., Fraioli,R. and Hood,L.E.
TITLE Construction of a Characterized Clone Resource for Genomic
Sequencing
JOURNAL Unpublished (1998)
COMMENT
Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Research Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 766 row: F column: 3
Seq primer: T7
Class: BAC ends
High quality sequence stop: 575.

```

```

FEATURES
source
1..575
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=766 Col=3 Row=F"
/clone_lib="RPCI-11 Human Male BAC Library"
/sex="male"
/note="Vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT 240 a 78 c 81 g 163 t 13 others
ORIGIN

```

```

alignment_scores:
Quality: 66.50 Length: 53
Ratio: 1.956 Gaps: 2
Percent Similarity: 64.151 Percent Identity: 28.302

```

```

alignment_block:
US-09-209-961-21 x A0620563/rev ..
Align seg 1/1 to reverse of: A0620563 from: 1 to: 575
4 lIelleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerG1 20
:|||||...:|||||...:|||||
247 ATGTTGCATAGACTTTTGTGTGTTATTTTATATATGTTCTATCTCC 198
20 uLeuTyrIleHisGlyTyrThrAspIleValLeuVal..... 33
|||||: ||| :|||:|||||:|||||:

```

```

197 TTTGTGG...CACATAAGTTCGACTTTTGTCTTAATCATCAATCAAT 151
34 .....ValPheThrValPheGluArgSerAlaGluAlaTyrValVal 47
:|||||...:|||||...:|||||
150 TTTGCCATATATTTGTGATTTTAAATTTCTTAAAGATGCTATTATCC 101
48 HisIleSer 50
|||||...:
100 CATGTAGCT 92
seq_name: gb_est24:A1226365
seq_documentation_block:
LOCUS A1226365 410 bp mRNA EST 29-OCT-1998
DEFINITION uf07b09.v1 Soares mouse mammary gland NMLMG Mus musculus cDNA clone
IMAGE:1510649 5', mRNA sequence.
ACCESSION A1226365
NID 93809418
VERSION A1226365.1 GI:3809418
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 410)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045169.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:937501
Seq primer: -40RP from Gibco
High quality sequence stop: 346.
FEATURES
source
1..410
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1510649"
/clone_lib="Soares mouse mammary gland NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT73 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT 77 a 99 c 116 g 118 t
ORIGIN

```

```

alignment_scores:
Quality: 63.50 Length: 51
Ratio: 2.048 Gaps: 1
Percent Similarity: 60.784 Percent Identity: 27.451
alignment_block:

```

```

US-09-209-961-21 x AI226365 ..
Align seg 1/1 to: AI226365 from: 1 to: 410

3 PheIleHisLeuGlyPheLysTrpClyValPheLysIleLysPheSe 19
|||||.....||| |||.....||| .....
167 TTTCTTATGCATCGGGAGACCCCTGGTGCTGTGGTCGTGCCGTCAG 216
rGlutLeuTyrlleHisGlyTyR.....ThrAspIleValV 31
|||||.....||| |||.....||| .....
217 TGAGATAFTTCACAGGCACTTGGCTTCGCCACAAGAAGTCCCTG 266
31 alLeuValValPheThrvAlpheGluArgSerAlaGluAlaTyrvAlval 47
|||||.....||| |||.....||| .....
267 TCGTAGTCAGTGCCTTATTGTTTTCTATTGAAGACCAGTTCGTTGCTTG 316

48 His 48
|||
317 CAC 319

seq_name: gb_est31:AU060237

seq documentation block:      697 bp   mRNA    EST       20-MAY-1999
LOCUS AU060237 Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLA658, mRNA sequence.
ACCESSION AU060237
NID 94881341
VERSION AU060237.1 GI:4881341
KEYWORDS EST.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum
REFERENCE Eukaryota; Dictyosteliida; Dictyostelium.
AUTHORS Yoshino,R., Morio,T. and Tanaka,Y.
TITLE Developmental CDNA in Dictyostelium discoideum
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3187168.

Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@sakura.cc.tsukuba.ac.jp
PROJECT = Dictyostelium discoideum cDNA project in Japan.

FEATURES             Location/Qualifiers
     source            1..697
                        /organism="Dictyostelium discoideum"
                        /strain="AX4"
                        /db_xref="taxon:44689"
                        /map="15 q15(21.1)-q22; 1: 21q"
                        /clone="SLA658"
                        /clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
                        /dev_stage="slug"
BASE COUNT          262 a 107 c 98 g 230 t
ORIGIN

alignment_scores:
Quality: 65.00           Length: 43
Ratio: 2.097             Gaps: 0
Percent Similarity: 72.093 Percent Identity: 30.233

alignment_block:
US-09-209-961-21 x AU060237/rev ..

Align seg 1/1 to reverse of: AU060237 from: 1 to: 697

7 LeuGlyPheLysTrpGlyValPheLysIleLysPheSerGluLeuTyrl 23
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
185 ATTGATTTCTCTGTTCTTATTAAATATCATTTTCTTCATTAATTT 136
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::
23 eHisGlyTyrrThrAsplelvalLeuValPheThrvAlpheGluA 40
::: |||::: |||::: |||::: |||::: |||::: |||::: |||:::

```

189 GTGCTCCCAATGACTTGCGAAAAATTTGCTCTTATCTACTACATTA 238

38 he.....GluArgSerAlaGluAlaTyrValVal 47

239 TTTAATATGAAGAAGGAGTGAATCTTATATGTT 273

seq_name: gb_gss10:AQ736850

seq_documentation_block: 438 bp DNA GSS 15-JUL-1999
 LOCUS AQ736850
 DEFINITION HS_2244_B2_H04_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate-2244 Col-8 Row-P, genomic survey sequence.
 ACCESSION AQ736850
 NID 95508402
 VERSION AQ736850.1 GI:5508402
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 438)
 AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J., Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
 TITLE Construction of a Characterized Clone Resource for Genomic Sequencing
 JOURNAL Unpublished (1998)
 COMMENT Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2244 row: P column: 8
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 438.

FEATURES

source
 1..438
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pBelobAC11; BAC Clones in E-Coli DH10B"
 BASE COUNT 102 a 100 c 88 g 146 t 2 others
 ORIGIN

alignment_scores:
 Quality: 63.00 Length: 53
 Ratio: 1.853 Gaps: 1
 Percent Similarity: 64.151 Percent Identity: 24.528

alignment_block:

US-09-209-961-21 x AQ736850

Align seg 1/1 to: AQ736850 from: 1 to: 438

1 MetValPheIleIleHisLeuGlyPheLysTyrGlyValPheLysIleLeu 17

265 CTATATTTTCTATCATCTTGTCATCTNTGGGAAGACAGATATA 314

17 sPheSerGluLeuTyrIleHisGlyTyrThrAspIleValValLeuVal 34

315 TTGTCCTCCAGCTCTCTCAGTAACCTC.....ATATTCATACATCCGA 358

34 aPheThrValPheGluArgSerIleGluAlaTyrValValHisIleSer 50

359 TTCTCTTATATTCAGAGGCACTCCAAGTGATGTAGTCTCCTGATG 408

51 ArgGlyLeu 53

409 AATGGAATA 417

seq_name: gb_gss7:AQ501328

seq_documentation_block: 639 bp DNA GSS 28-APR-1999
 LOCUS AQ501328
 DEFINITION V24C4 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae genomic 5', genomic survey sequence.
 ACCESSION AQ501328
 NID 94706978
 VERSION AQ501328.1 GI:4706978
 KEYWORDS GSS.
 SOURCE baker's yeast.
 ORGANISM Saccharomyces cerevisiae
 Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 REFERENCE 1 (bases 1 to 639)
 AUTHORS Ross-Macdonald,P., Roemer,T., Coelho,P.S.R., Agarwal,S., Kumar,A., desEtages,S.A., Cheung,K.-H., Sheehan,A., Symoniatidis,D., Jansen,R., Umansky,L., Heitman,M., Nelson,K., Iwasaki,H., Kanada,D., Lugo,R., Hager,K., Miller,P., Roeder,G.S. and Snyder,M.
 TITLE Large-Scale Analysis of the Yeast Genome by Transposon Tagging and Gene Disruption
 JOURNAL Unpublished (1999)
 COMMENT Contact: Kumar A
 Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
 Yale University
 P.O. Box 208103, New Haven, CT 06520-8103, USA
 Tel: 203 432 9949
 Fax: 203 432 6161
 Email: anuj.kumar@yale.edu
 te of mTn-3xHA/lacZ insertion.
 Seq primer: GGCTTCTTCTTTGGAGTAC
 Class: transposon-tagged.
 Location/Qualifiers
 1..639
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library"
 /lab_host="E. coli"

FEATURES

source
 1..639
 /organism="Saccharomyces cerevisiae"
 /db_xref="taxon:4932"
 /clone_lib="mTn-3xHA/lacZ Insertion Library"
 /note="Vector: pHS6-Sal; A yeast genomic DNA library (lacking mitochondrial DNA) was prepared in pHS6-Sal; genomic DNA was size-fractionated (DNA of roughly 2-3 kb in length) prior to cloning. This library was subsequently mutagenized with a mTn-3xHA/lacZ minitransposon containing lacZ, URA3, and tet resistance."
 BASE COUNT 181 a 137 c 141 g 179 t 1 others
 ORIGIN

alignment_scores:
 Quality: 64.00 Length: 54
 Ratio: 1.939 Gaps: 3
 Percent Similarity: 61.111 Percent Identity: 35.185

alignment_block:

US-09-209-961-21 x AQ501328/rev

Align seg 1/1 to reverse of: AQ501328 from: 1 to: 639

6 HisLeuGlyPheLys.TyrGlyValPheLysIleLysPheSerGluLeu 22

493 CACTTGGGCTTTTTCGGCAGGTGTTGCTCTCTATCTCTTTCAGTT 444

22 yrIleHisGlyTyrThrAspIleValValLeuValPheThrValPhe 38

443 TGGTCCGC...TATTCAGATTATTGGTCCTCTACTGATATCATCTCT 397

```
39 GluArgSer.....AlaGluAlaTyrValValHi 48
   ::::
   396 TCATTTCTTTCTTCACAGTGGGTAAATGGGAATACATACCA 347
       48 sileSerArg 51
       ::::
       346 TATCTCAGCT 337

seq_name: gb_gss8:AQ381971

seq_documentation_block:
LOCUS   AQ381971      656 bp      DNA      20-MAY-1999
DEFINITION   RPC11-167M1-TV RPCI-11 Homo sapiens genomic clone RPCI-11-167M1,
genomic survey sequence.
ACCESSION   AQ381971
NID        94352994
VERSION    AQ381971.1 GI:4352994
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 656)
AUTHORS    Zhao,S., Adams,M.D., Nierman,W., Malek,J., de Jong,P. and
Venter,J.C.
TITLE      Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL    Unpublished (1997)
COMMENT    Other_GSSs: RPC11-167M1.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department Of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbestigr.org
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from
Research Genetics (info@resgen.com). BAC end search page:
http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
Seq primer: T7
Class: BAC ends.
FEATURES             Location/Qualifiers
     source           1..656
                     /organism="Homo sapiens"
                     /db_xref="GB:7564032"
                     /db_xref="taxon:9606"
                     /clone="RPCI-11-167M1"
                     /clone_lib="RPCI-11"
                     /sex="Male"
                     /cell_type="Lymphocytes"
                     /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT        168 a   142 c   103 g   242 t
ORIGIN
```

```
alignment_scores:
  Quality: 64.00      Length: 44
  Ratio: 2.000       Gaps: 2
  Percent Similarity: 72.727  Percent Identity: 31.818

alignment_block:
US-09-209-961-21 x AQ381971  ..
Align seg 1/1 to: AQ381971  from: 1 to: 656

4 lilelleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerC1 20
:::
60 CTCCTGCCAATGGCCCTTTTGTGGGGGCTCATGTTCCTCTCTCTGTC 109
```

```
20 uleuTyrrileHisGlyTyrrThraspile...ValValLeuValValPhe 36
   ::::
   110 CTAATTTTATCTCGTACACTGCAGTCAATGTTAGCACACATCTTATTTT 159
       36 hrValPheGluArgSerAlaGluAlaTyrVal 46
       ::::
       160 CATTATTGAA...GCTGTCCAAGCATTCATA 188

seq_name: gb_gss6:AQ410083

seq_documentation_block:
LOCUS   AQ410083      473 bp      DNA      16-MAR-1999
DEFINITION   HS.5100.B2.E04.SP6E RPCI-11 Human Male BAC Library Homo sapiens
genomic clone Plate-676 Col-8 Row-J, genomic survey sequence.
ACCESSION   AQ410083
NID        94432067
VERSION    AQ410083.1 GI:4432067
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 473)
AUTHORS    Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Furlong,J.,
Shaker,R., Schmidt,S., Traicoff,R. and Hood,L.E.
TITLE      Construction of a Characterized Clone Resource for Genomic
Sequencing
JOURNAL    Unpublished (1998)
COMMENT    Contact: Mahairas GG, Wallace JC, Hood L
High Throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones are derived from the human BAC library RPCI-11. For BAC
library availability, please contact Pieter de Jong
(pieterejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
or from Resear h Genetics (info@resgen.com). BAC end Web Server:
http://www.htsc.washington.edu
Plate: 676 row: J column: 8
Seq primer: SP6
Class: BAC ends
High quality sequence stop: 473.
FEATURES             Location/Qualifiers
     source           1..473
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="Plate-676 Col-8 Row-J"
                     /clone_lib="RPCI-11 Human Male BAC Library"
                     /sex="male"
                     /note="vector: pBACe3.6; Genomic sequence of BAC ends"
BASE COUNT        220 a   85 c   62 g   105 t
ORIGIN
```

```
alignment_scores:
  Quality: 62.00      Length: 36
  Ratio: 2.480       Gaps: 1
  Percent Similarity: 69.444  Percent Identity: 38.889

alignment_block:
US-09-209-961-21 x AQ410083/rev  ..
Align seg 1/1 to reverse of: AQ410083  from: 1 to: 473

2 ValPheIleIleHisLeuGlyPheLysTrpGlyValPheLysIleLysph 18
:::
164 CTATTTGCCAATCTATATCTTTTAAATGGGCATTAATTCATTTATTT 115
```


224 CTATTC 229

seq_name: gb_est25:AI320780

seq_documentation_block: 370 bp mRNA EST 18-DEC-1998
LOCUS AI320780
DEFINITION d1a12m.r1 Neurospora crassa morning cDNA library Neurospora crassa
cDNA clone d1a12m 3', mRNA sequence.

ACCESSION AI320780

NID 94036762

VERSION AI320780.1 GI:4036762

KEYWORDS EST.

SOURCE Neurospora crassa.

ORGANISM

Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;

Sordariales; Sordariaceae; Neurospora.

1 (bases 1 to 370)

Zhu, H., Lai, H., Kupfer, D., Dunlap, J. C. and Roe, B. A.

Two Neurospora crassa EST Databases

Unpublished (1998)

On Oct 19, 1998 this sequence version replaced gi:454479.

Other ESTs: d1a12m.fl

Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu

Department of Chemistry and Biochemistry

Advanced Center for Genome Technology, University of Oklahoma

620 Parrington Oval, Norman, OK 73019, USA

Tel: 405 325 4912

Fax: 405 325 7762

Email: broe@ou.edu

We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center

Possible reversed clone: polyT not found

Seq primer: Universal Reverse Primer

High quality sequence stop: 323.

FEATURES

source

1..370

/organism="Neurospora crassa"

/strain="bd, frq7 A"

/db_xref="taxon:5141"

/clone="d1a12m"

/clone_lib="Neurospora crassa morning cDNA library"

/tissue_type="tissue harvested following 22hr growth in

dark"

/notes="Vector: pBluescript SK-; Site_1: XbaI; Site_2:

EcoRI; See: Bell-Pedersen, D., et al. PNAS 93:13096,1996.

5' end of cDNA cloned into XbaI site of pBluescript; 3'

end of cDNA cloned into EcoRI site of pBluescript"

BASE COUNT 113 a 104 c 91 g 62 t

ORIGIN

alignment_scores:

Quality: 60.50 Length: 69

Ratio: 1.729 Gaps: 3

Percent Similarity: 50.725 Percent Identity: 26.087

alignment_block:

US-09-209-961-21 x AI320780/rev ..

Align seg 1/1 to reverse of: AI320780 from: 1 to: 370

3 PheIleHisLeuGlyPheLysTrp..... 11

|||||..... 11

254 TTCATCTCCGGTTTGCCTTCAATGGGTGGGATTTTGCTCTTTAT 205

12 GlyValPheLysIleLysPheSerCluLeuTrpIleHisG 25

|||||..... 25

204 TCTTTGGTCCCGCTTCCTCGCGGTC...TTCCGGACATGATACATCCATG 158

|||||..... 158

25 ly..... 26

|||||..... 26

157 AGAGCTTGAGGGGCACACACGATATTCCTCGATGAACATGCCCTCGG 108

|||||..... 108

27 ThrAspIleValLeuValValPheThrValPheGluArgSerAlaG1 43

|||||..... 43

107 GTGACCTCGTCAACATGCTTCTGGCTCGTCAATCGCGTTTCTGCATT 58

|||||..... 58

43 uLaTyr 45

|||||

57 GGCCTTAC 51

seq_name: gb_gss5:AQ331923

seq_documentation_block:

LOCUS AQ331923 378 bp DNA GSS

DEFINITION HS_5002_A2_E08_SP6E RPC111 Human Male BAC Library Homo sapiens

genomic clone Plate=578 Col=16 Row=I, genomic survey sequence.

ACCESSION AQ331923

NID 94128510

VERSION AQ331923.1 GI:4128510

KEYWORDS GSS.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 378)

Mahairas, G. G., Wallace, J. C., Smith, K., Swartzell, S., Furlong, J.,

Shaker, R., Schmidt, S., Traicoff, R. and Hood, L. E.

Construction of a Characterized Clone Resource for Genomic

Sequencing

Unpublished (1998)

JOURNAL

COMMENT

Contact: Mahairas GG, Wallace JC, Hood L

High Throughput Sequencing Center

University of Washington

401 Queen Anne Avenue North, Seattle, WA 98109, USA

Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones may be purchased from Research Genetics (info@resgen.com).

BAC end Web Server: http://www.htsc.washington.edu

Plate: 578 row: I column: 16

Seq primer: SP6

Class: BAC ends

High quality sequence stop: 378.

FEATURES

source

1..378

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="plate=578 Col=16 Row=I"

/clone_lib="RPC111 Human Male BAC Library"

/sex="Male"

/cell_type="Lymphocytes"

/note="vector: pBAC3.6; RPC111 Human Male BAC Library"

BASE COUNT 144 a 50 c 47 g 137 t

ORIGIN

alignment_scores:

Quality: 60.50 Length: 45

Ratio: 2.327 Gaps: 1

Percent Similarity: 57.778 Percent Identity: 26.667

alignment_block:

US-09-209-961-21 x AQ331923 ..

Align seg 1/1 to: AQ331923 from: 1 to: 378

5 IleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerCluLe 21

|||||..... 21

242 GTGCCTAACCTTTCTCTCTATTTTCTCAATTAATGATGTCAGAT 291

|||||..... 291

21 utyrtileHisGlyTyrtThrAspIleValLeuValPheThrValp 38

|||||..... 38

292 ACAGACACACACACACGATATATATATATAGTA..... 328

|||||..... 328

```

38 heGluArgSerAlaGluAlaTyrValValHisIle 49
|||||:|||||
329 .....TATGTACTATATATT 343

seq_name: gb_est35:AV158479

seq_documentation_block:
LOCUS AV158479 180 bp mRNA EST 07-JUL-1999
DEFINITION AV158479 Mus musculus head C57BL/6J 12-day embryo Mus musculus cDNA
clone 3010016C09, mRNA sequence.
ACCESSION AV158479
NID 95402114
VERSION AV158479.1 GI:5402114
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 180)
AUTHORS Carninci,P., Shibata,K., Ozawa,Y., Konno,H., Itoh,M., Aizawa,K.,
Akahira,S., Akiyama,J., Fukuda,S., Fukunishi,Y., Funayama,T.,
Hara,A., Hayatsu,N., Hori,F., Ishikawa,T., Itoh,M., Izawa,M.,
Kawai,J., Kikuchi,N., Kojima,Y., Matsuyama,T., NiiTsuna,H., Oda,H.,
Owa,C., Sato,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y.,
Sugahara,Y., Suzuki,H., Suzuki,H., Tateno,M., Tomaru,Y.,
Tomimaga,N., Watanabe,S., Yagame,M., Yamamura,T., Yokota,T.,
Yoshino,M., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.
Riken Mouse ESTs
JOURNAL Unpublished (1999)
COMMENT On May 18, 1998 this sequence version replaced gi:3138222.

Contact: Chie Owa
Genome Science Laboratory
RIKEN
3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan
Tel: 81-298-36-9145
Fax: 81-298-36-9098
Email: genome-res@r.riken.go.jp
Thermostabilization and thermoactivation of thermolabile enzymes by
trehalose and its application for the synthesis of full length cDNA
(Proc. Natl. Acad. Sci. U.S.A. 95(2):520-524 (1998))
Transcriptional sequencing: A method for DNA sequencing using RNA
polymerase (Proc. Natl. Acad. Sci. U.S.A. 95(7):3455-3460 (1998))
Please visit our web site (http://genome.rtc.riken.go.jp) for
further details.
FEATURES
source
location/Qualifiers
1..180
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="3010016C09"
/clone_lib="Mus musculus head C57BL/6J 12-day embryo"
/sex="mixed"
/tissue_type="head"
/dev_stage="12-day embryo"
BASE COUNT 44 a 49 c 37 g 50 t
ORIGIN

alignment_scores:
Quality: 58.00 Length: 47
Ratio: 2.071 Gaps: 0
Percent Similarity: 59.574 Percent Identity: 36.170

alignment_block:
US-09-209-961-21 x AV158479 ..
Align seg 1/1 to AV158479 from: 1 to: 180

5 IleHisLeuGlyPheLysTrpGlyValPheLysIleLysPheSerGluLe 21
|||||:|||||
30 ATTAACCTTGCGGATAAATGGGGGTAGGGGATATATTCACACTTTGGG 79
.....

seq_name: gb_est25:AI319299

seq_documentation_block:
LOCUS AI319299 402 bp mRNA EST 18-DEC-1998
DEFINITION a8902nm.r1 Neurospora crassa morning cDNA library Neurospora crassa
cDNA clone a8902nm 3', mRNA sequence.
ACCESSION AI319299
NID 94035281
VERSION AI319299.1 GI:4035281
KEYWORDS EST.
SOURCE Neurospora crassa.
ORGANISM Neurospora crassa
Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
Sordariales; Sordariaceae; Neurospora.
REFERENCE 1 (bases 1 to 402)
AUTHORS Zhu,H., Lai,H., Kupfer,D., Dunlap,J.C. and Roe,B.A.
TITLE Two Neurospora crassa EST Databases
JOURNAL Unpublished (1998)
COMMENT On Aug 21, 1998 this sequence version replaced.
Other_ESTs: a8902nm.fl
Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
Department of Chemistry and Biochemistry
Advanced Center for Genome Technology, University of Oklahoma
620 Parrington Oval, Norman, OK 73019, USA
Tel: 405 325 4912
Fax: 405 325 7762
Email: broe@ou.edu
We anticipate the future release of the cDNA clones to the Fungal
Genetics Stock Center
Possible reversed clone: polyt not found
Seq primer: Universal Reverse Primer
High quality sequence stop: 322.
FEATURES
source
location/Qualifiers
1..402
/organism="Neurospora crassa"
/strain="bd, fig7 A"
/db_xref="taxon:5141"
/clone="a8902nm"
/clone_lib="Neurospora crassa morning cDNA library"
/tissue_type="tissue harvested following 22hr growth in
dark"
/notes="Vector: pBlueScript SK-; Site 1: XbaI; Site 2:
EcoRI; See: Beil-Pedersen,D., et al. PNAS 93:13095,1996.
5' end of cDNA cloned into XbaI site of pBluescript; 3'
end of cDNA cloned into EcoRI site of pBluescript"
BASE COUNT 128 a 112 c 99 g 63 t
ORIGIN

alignment_scores:
Quality: 60.50 Length: 69
Ratio: 1.729 Gaps: 3
Percent Similarity: 50.725 Percent Identity: 26.087

alignment_block:
US-09-209-961-21 x AI319299/rev ..
Align seg 1/1 to reverse of: AI319299 from: 1 to: 402

3 PheIleIleHisLeuGlyPheLysTrp..... 11
|||||:|||||
254 TTCATCTCGGTTTGCTTCAATGTTGGGGATTTTGCTCTTTAT 205
.....

12 .....GlyValPheLysIleLysPheSerGluLeuTyrIleHisG 25
|||||:|||||

```

```
204 TCTTTGGTCGCCGCTCTTCGGGGTC...TTCGGGGACATGTACATCCATG 158
25 ly.....Tyr 26
157 AGAGCCTGAGGGCAACAACGATATTCGTCGGATGAAGCATGCCCGTCTGG 108
27 ThrAspIleValValLeuValPheThrValPheGluArgSerAlaG1 43
107 GTGGACCTCGTCAACATGCTTCTCTGGCTCGTCATCGCGGTTTCTGCATT 58
43 ualATyr 45
57 GGCCTAC 51
```


OM of: US-09-209-961-3 to: GenEmbl.* out_format : pfs

Date: Dec 27, 1999 2:12 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+p2n.model1 -DEV=xlp
-O=/cgn2.1/USPTO_spool/US09209961/runat_22121999_101043_17910/app_query.fasta.1
-DB=GenEmbl -QWMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=score -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-3
Query length: 314
Database: GenEmbl.*
Database sequences: 780561
Database length: 2137953050
Search time (sec): 1994.760000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_vi:AF085695	+ 1714.00	2896.36	2.5e-153	1768	! AF085695 Porcine circovirus st
gb_vi:AF086836	+ 1710.00	2889.55	6.0e-153	1768	! AF086836 Porcine circovirus st
gb_vi:AF086834	+ 1706.00	2882.74	1.4e-152	1768	! AF086834 Porcine circovirus st
gb_vi:AF086835	+ 1701.00	2874.24	4.3e-152	1768	! AF086835 Porcine circovirus st
gb_vi:AF055391	+ 1682.00	2841.90	2.7e-150	1768	! AF055391 Porcine circovirus ty
gb_vi:AF055392	+ 1682.00	2841.90	2.7e-150	1768	! AF055392 Porcine circovirus ty
gb_vi:AF053937	+ 1682.00	2841.90	2.7e-150	1768	! AF053937 Bovine circovirus, co
gb_vi:AF147751	+ 1682.00	2841.90	2.7e-150	1768	! AF147751 Porcine circovirus, ty
gb_vi:PCJ34185	+ 1682.00	2841.90	2.7e-150	1768	! AJ223185 Porcine circovirus DN
gb_vi:AF166528	+ 1678.00	2835.09	6.5e-150	1768	! AF166528 Porcine circovirus co
gb_vi:AF109399	+ 1677.00	2833.39	8.1e-150	1768	! AF109399 Porcine circovirus ty
gb_vi:AF055393	+ 1676.00	2831.70	1.0e-149	1767	! AF055393 Porcine circovirus ty
gb_vi:AF055394	+ 1676.00	2831.70	1.0e-149	1767	! AF055394 Porcine circovirus ty
gb_vi:AF027217	+ 1674.00	2831.69	1.0e-149	1768	! AF027217 Porcine circovirus st
gb_vi:AF117753	+ 1674.00	2828.29	1.6e-149	1768	! AF117753 Porcine circovirus ty
gb_vi:AF112862	+ 1671.00	2823.18	3.0e-149	1768	! AF112862 Porcine circovirus ty
gb_vi:AF154679	+ 1669.00	2819.78	4.6e-149	1768	! AF154679 Porcine circovirus, c
gb_vi:AF109398	+ 1665.00	2812.97	1.1e-148	1768	! AF109398 Porcine circovirus ty
gb_vi:AF012107	+ 1447.00	2442.03	5.1e-128	1759	! AF012107 Porcine circovirus, c
gb_vi:PCU049186	+ 1447.00	2442.03	5.1e-128	1759	! U49186 Porcine circovirus, com
gb_vi:AF071879	+ 1443.00	2435.23	1.2e-127	1758	! AF071879 Porcine circovirus, c
gb_vi:PCCOMGEN	+ 1439.00	2428.41	2.9e-127	1758	! Y09921 Porcine circovirus comp
gb_vi:AF118095	+ 1063.00	1799.10	3.3e-92	597	! AF118095 Porcine circovirus ty
gb_vi:AF080560	+ 610.50	1017.25	1.2e-48	1993	! AF080560 Beak and feather dise
gb_vi:AF071878	+ 597.50	995.13	2.0e-47	1993	! AF071878 Beak and feather dise
gb_vi:AF135393	+ 289.00	484.94	5.2e-19	438	! AF135393 Porcine circovirus PCV
gb_vi:AF134203	+ 261.00	437.33	2.3e-16	436	! AF134203 Porcine circovirus PCV
gb_vi:BBU12586	+ 197.00	319.27	8.9e-10	1111	! U12586 Banana bunchy top virus
gb_vi:BBU12587	+ 197.00	319.13	8.9e-10	1127	! U12587 Banana bunchy top virus
gb_vi:NYV5968	+ 185.50	300.69	9.5e-09	1004	! AJ005968 faba bean necrotic ye
gb_vi:SCU01731	+ 182.00	294.56	2.1e-08	1022	! U16731 Subterranean clover stu
gb_vi:ABU000921	+ 168.50	271.71	3.9e-07	1009	! AB000921 Milk vetch dwarf viru
gb_pa:AR063451	+ 168.50	270.82	4.4e-07	1106	! AR063451 Sequence 3 from paten
gb_vi:BYTV1	+ 168.50	270.82	4.4e-07	1106	! L32166 Banana bunchy top virus
gb_vi:CFDCG	+ 163.00	266.75	7.4e-07	1291	! M2963 Coconut foliar decay vi
gb_pa:AR063452	+ 163.00	261.55	1.4e-06	1096	! AR063452 Sequence 4 from paten
gb_vi:BYTV2	+ 163.00	261.55	1.4e-06	1096	! L32167 Banana bunchy top virus
gb_vi:U97525	+ 157.00	249.127	5.4e-06	1103	! U97525 Banana bunchy top virus
gb_pa:AR063453	+ 156.00	249.68	6.7e-06	1091	! AR063453 Sequence 5 from paten
gb_pa:AR010234	+ 156.00	249.55	6.7e-06	1105	! AR010234 Sequence 18 from paten
gb_vi:U02312	+ 152.50	243.69	1.4e-05	1095	! U02312 Banana bunchy top virus
gb_pa:AR010232	+ 152.00	242.76	1.6e-05	1103	! AR010232 Sequence 16 from paten

gb_pat:AR010233 + 152.00 242.75 1.6e-05 1104 ! AR010233 Sequence 17 from p
gb_vi:AF102780 + 150.50 240.14 2.2e-05 1111 ! AF102780 Banana bunchy top
gb_pat:AR010225 + 147.50 235.04 4.3e-05 1111 ! AR010225 Sequence 9 from pa

seq_name: gb_vi:AF085695

seq_documentation_block:

LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998
DEFINITION Porcine circovirus strain 412, complete genome.
ACCESSION AF085695
NID 93688362
VERSION 1 GI:3668362
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.

TITLE Direct Submission

JOURNAL Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,
University of Saskatchewan, 120 Veterinary Road, Saskatoon,
Saskatchewan S7N 5E3, Canada

FEATURES

source Location/Qualifiers

1..1768

/organism="porcine circovirus"

/strain="412"

/db_xref="taxon:46221"

51..995

/codon_start=1

/product="P3.8"

/protein_id="AAC61860.1"

/db_xref="PID:g3668363"

/db_xref="GI:3668363"

/translation="MPSKNGRSQPQPKFWVTLNPNSEDERKKIRELPISLDFYFI
VEEGNEERTPHQGFANFVKQTNKVWYLGARCHIEKAKGTDOQNEKCYSGEN
LIECKATNPVSGDLSTAVSLLEGILTVAKQHPVTFVFNFRGLAEKLVSGKM
QRDMTGVHFIIVGPGCKSKWAFANPETTYWPKPKMKWDGYGKGVVDDFY
GWLPMDDLRLCDRYPLTVKTKGTVPFLARSLITNSQPLEWYSSTAVPAVALYR
RTLSVFNKAKQSTEEGQFVTLSPCPPEFFEYNY"
complement(357..671)

/codon_start=1

/product="P12.1"

/protein_id="AAC61865.1"

/db_xref="PID:g3668368"

/db_xref="GI:3668368"

/translation="WTIIPLVRFNFPVCGFRVCKISSPFAFTTPRPHNEVYIGFPI
TLHFFPAHFQKFSQPAIEFDKRYRLVLLCNQHONPALQOQGHSSRQVTPLSRSSTF
NK"
complement(386..565)

/codon_start=1

/product="P6.6"

/protein_id="AAC61866.1"

/db_xref="PID:g3668369"

/db_xref="GI:3668369"

/translation="MKTLLVFQSRFCIFLTFKSSASPRKFLTNVTGCCFATVTRPL
SNKVLTAVDRLRCP"
553..732

/codon_start=1

/product="P6.6"

/protein_id="AAC61861.1"

/db_xref="PID:g3668364"

/db_xref="GI:3668364"

/translation="MTLSMGLGVVKGALLILOTRKPHTGHLKTSGGWTVWKKW
LILMTFMACRGMHY"
833..970

/codon_start=1

/product="P5.0"

/protein_id="AAC61862.1"

/db_xref="PID:g3668365"

/db_xref="GI:3668365"
/translation="MVLNCCPSCRSLSIEDYFLGLEECYKTIHGGRGVRHPPPPM
P"

CDS

1016 ..1177

/codon_start=1
/product="P6.2"
/protein_id="AAC61863.1"
/db_xref="PID:g3668366"
/db_xref="GI:3668366"
/translation="MVFIIHLGFKMGVFKIFSELYIHGYTDIVLVVFTVFSAEAE
YVHISKGL"

CDS

complement(1034..1735)

/codon_start=1
/product="P27.9"
/protein_id="AAC61864.1"
/db_xref="PID:g3668367"
/db_xref="GI:3668367"
/translation="MTIPRRYRRRRHRRPSHLGQILRRRPLVHPHRRYRWRKNGI
FNTLSRTFGYTVKRTVTTPSWAVDMRMFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTIDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLSTIDYFQPNKRNQLRLQTSNVVDHVLGLTAFENSKYDODYNIWTMY
VQFEFNKDPLEP"

BASE COUNT 463 a 362 c 481 g 462 t

ORIGIN

alignment_scores:

Quality: 1714.00 Length: 314
Ratio: 5.459 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.682

alignment_block:

US-09-209-961-3 x AF085695

Align seq 1/1 to: AF085695 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCGAGCAAGAAGATGGAAGAGCGGACCCCAACCATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
|||||
101 GGTGTTCACGCTGAATAATCTTCCGAGCAGCAGCAAGAAAATACGGG 150
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
151 AGCTCCCAATCTCCCTATTGATATTATTGTTGGCGAGGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAGGAGCAGACACTCACCCTCAGGGGTTGCTAATTTCTGAGAA 250
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84
|||||
251 GCAAACTTTTAAATAGTGAAGTGTATTGTTGGTCCCGCTGCCACATCG 300
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCCAAAGAACTGATCAGCAAGTAATAAGAAATATTGCAGTAAGA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GGCACCTTACTTATTGAATGTGAGTCCCTCCGATCTCAAGACACGAG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTGTCTACTGCTGAGTACCTTGTGGAGAGCGGATCTCGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCCTTGCAAGCAGCAGCCCTGTAACTGTTTCACAAAATTTCCCGCGGCTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167

|||||
501 GCTGAACATTTGAAGTGACGGGAAAATGAAAAGCGTGATTGGAAC 550
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
551 CAATGTACACTTCATTGTGGGGCCACCTGGGTGTGTAAGCAATGGG 600
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAATTTTGCACCCCGAACCACACATACTGGAACCCACCTAAAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217
|||||
651 AAGTGTGGGATGTTACCATGGTGAAGAAAGTGGTCTTATTGATGACTT 700
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGCTCCGCTGGGATGATCTACTGAGACTGTGATCGATATC 750
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
751 CATTGACTGTAAAACATAAGGTGGAACTGTACTCTTTTGGCCCGCAGT 800
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCTGATTACCAGCAATCAGACCCCGTTGGAATGGTACTCCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCGCCAGCTGTAGAGCTCTCTATCGGAGGATTACTTCTTGGTATTT 900
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
|||||
901 GGAAGATGCTACAAACAATCCAGGAGGAGGGGCCAGTTCGTCACC 950
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCATCGCCCTGAATTTCCATATGAAATAAATTAC 992

seq_name: gb_v1:AF086836

seq_documentation_block:

LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain M226, complete genome.

ACCESSION AF086836

NID 93661521

VERSION AF086836.1 GI:3661521

KEYWORDS

porcine circovirus.

SOURCE

Porcine circovirus

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES

source

1..1768 Location/Qualifiers

/organism="porcine circovirus"

/strain="M226"

/db_xref="taxon:46221"

51..995

/codon_start=1

/product="P35.8"

/protein_id="AAC61741.1"

/db_xref="PID:g3661522"

/db_xref="GI:3661522"

/translation="WPSKKNRSGPQPKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGSRTPHLOGFANFYKQTFNKVKWYFGARCHIEKAKGTDOONKEYCSKEGN

LLIECGAPRSQORSDLSTAVSTLLESGILVTVAKQHPVTVFNFRGLAELLKVSGRM
QKRDWNTNHFIVGPPCGSKSWANFANPETTYWRPNKNKWDGYHGEKVWVDDFY
GTLPWDDLRLCDRYPLTVKTKGCTVPLFARSILITSNQTPLEWYSTAVPAVEALYR
RTSLVFWKNATKQSTEEGQFVTLSPPCPEFFPEINY
complement(1034..1735)

CDS

/codon_start=1
/product="P27.9"
/protein_id="AAC61742.1"
/db_xref="PID:g3661523"
/db_xref="GI:3661523"
/translation="MYPRRYRRRPRSHLQILRRPWLVRHRYRWRKNGI
FNRLSKTGITVKRTVTTPSWADMRFKIDDFVPPGGTNRKISIPFFYIRKVK
VEWPCSPITQGRGVSFAVILDDNFVKTATLYDPYVNSRHTIPQPFYHSRY
FTPKPVLDSITDYFOENNRNQLWLRLQTSNVDHVLGLTAFENSKYDQDYNIRVTMY
VQPRENLKDPPLP"

BASE COUNT 463 a 363 c 480 g 462 t
ORIGIN

alignment_scores:

Quality: 1710.00 Length: 314
Ratio: 5.446 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.363

alignment_block:

US-09-209-961-3 x AF086836 ..

Align seg 1/1 to: AF086836 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCACGACGAGAGAAATGGAAGAGCGGCCCAACACACATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
101 GGTGTTCACGCTCATATACCTTCGAGACGAGCGCAAGAAAATACGGG 150
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
151 AGCTCCCAATCTCCCTATTATTGATTATTTTGTGGCAGAGGGGTAA 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
201 GAGGAGCGCCGACACACCCACCTACAGGGGTTCGCTAATTTGTGAAGA 250
67 sGlnThrPheAsnLysValLysTyrTrpLeuGlyAlaArgCysHisIleG 84
251 GCAAACTTTTAATAAGTGAAGTGGTATTTTGGTCCCGCTGCCACATCG 300
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
301 AGAAGCGAAAGGAACTGATCAGCAGATAAAGATAATTCAGTAAAGAA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
351 GGCACACTTACTATTGATGATGGAGCTCTCGATCTCAAGGACACGGAG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
401 TGACCTGTCTACTGCTGTGAGTACCTTCTGGAGAGCGGGATTCGTGTA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCGTGTGCAAGACGACGCCCTGTAAGCTTGTCAAAAATTTCCCGGGCTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
501 GCTGAACCTTTTGAAGTACGCGGGAATGCAAAAGCGGTGATTGGAAGAA 550
167 rAsnValHisPheIleValGlyProGlnGlyCysGlyLysSerLysTyr 184
551 CAATGTACACTTCATTGTGGGGCCACCTGGGTGTTGTTAAAGCAATGGG 600

184 laaLaasnPheAlaasnProGluThrThrTyrTrpLysProProLysAsn 200
601 CTGCTAAATTTTCAAAACCGGAAACACACTACTGGAACACCACTAAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
651 AAGTGGTGGGATGGTTACCATGGTGAAGAAAGTGGTTGTTATTGATGACT 700
217 eYrGlyTrpLeuProTyrAspAspLeuLeuArgLeuCysAspArgTyrP 234
701 TTATGGCTGGCTGGCTGGATGATCTACTGAGACTGTGTGATCGATATC 750
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
751 CATTCGCTGATAAACTAAAGTGGAACTGTACCTTTTGGCCCGCAGT 800
251 lleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
801 ATTCGATTACCAAGCAATCAGACCCCGTTGGAATGTACTCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
851 TGTCCGAGCTGTAGAAGCTCTCTATCGGAGGATTAATCTCTTGGTATTT 900
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
901 GGAAGAATGCTACAAACAATCCACGGAGAGGGGGCCGACCTCGTCACC 950
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
951 CTTTCCCCCATGCTGCTGAATTCATATGAAATAAATTAC 992
seq_name: gb_vi:AF086834
seq_documentation_block:
LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.
ACCESSION AF086834
NID 93661515
VERSION AF086834.1 GI:3661515
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus
REFERENCE 1 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:g3661516"
/db_xref="GI:3661516"
/translation="MPSKKNRSGPQPKRWFTLNNPSEDERKKIRELPISLDYFI
VGEENEECPHLOGFANFVKQTNKVKYLCARCHIEKAKCTDOONKEYCSEGN
LLIECGAPRSQORSDLSTAVSTLLESGILVTVAKQHPVTVFNFRGLAELLKVSGRM
QKRDWNTNHFIVGPPCGSKSWANFANPETTYWRPNKNKWDGYHGEKVWVDDFY
GTLPWDDLRLCDRYPLTVKTKGCTVPLFARSILITSNQTPLEWYSTAVPAVEALYR
RTSLVFWKNATKQSTEEGQFVTLSPPCPEFFPEINY"
complement(1034..1735)
/codon_start=1
/product="P27.9"

```
/protein_id="AAC61738.1"
/db_xref="PID:g3661517"
/db_xref="GI:3661517"
/translation="MTYPRRRYRRRRHRRPSHLGOILRRRPMWLVHPRHRYRWRKNGI
ENRLSRFTGTVKATTTTTPSWADMMRFNDIDFVPPGGGTNKISIPFEYIRKVK
VEFWPCSPITQDROGVSTAVILDDNFVKATLYIDPVNFYSSRHTIPOPFSYHSRY
FTPKPVLDSITDYFQPNKRNQLWLRQTSRNVHVGGLTAFENSKYDQDYNIRVTM
VQRFENLKDPLKP"
BASE COUNT      463 a 357 c 482 g 466 t
ORIGIN

alignment_scores:
  Quality: 1706.00      Length: 314
  Ratio: 5.433          Gaps: 0
Percent Similarity: 100.000 Percent Identity: 99.045

alignment_block:
US-09-209-961-3 x AF086834
Align seg 1/1 to: AF086834 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCAAGCAAGAAATGGAAGAAGCGGACCCCAACACATRAAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLysLeu 34
|||||
101 GGTGTCACCTGAATATCTCCGAAGACGAGCGCAAGAAATACGGG 150
|||||
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
151 AGCTCCCAATCCCTATTGATTATTTATTTGTCGAGGAGGGTAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
|||||
201 GAGGAGGAGCAACACCTCACCTCCAGGGTTCCTAATTTGTGAAGAA 250
|||||
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIle 84
|||||
251 GCAAACTTTTAAAGTAGAGTGGTATTGGGGTCCCGCTGCCACATCG 300
|||||
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAGCCAAAGGACGTATCAGCAGATAAAGATATTGTAGTAAGAA 350
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
|||||
351 GGCAACTTACTTATTGAATGTGGAGCTCCTCGATCTCAAGACACACGGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
|||||
401 TGACCTGCTACTGCTGTAGTACTCTGTGGAGAGCGGATCTGTGTA 450
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTGTGAAGACGACCCCTGTACAGTTTGTCAAAAATTTCCGGCGGCTG 500
|||||
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTGAACCTTTTGAAGTAGAGCGGAAAAATGCAAAAGCGTGATTGGAACA 550
|||||
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrp 184
|||||
551 CAATGTACATTCATTGTGGGGCCACCTGGGTGGTAAAGCAAAATGGG 600
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAATTTTGAACCCCGAAACCACTACTGGAACCAACCTCAAAAC 650
|||||
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217
|||||
651 AAGTGGTGGGATGGTTACCATGGTGAAGAAAGTGGTGTATTGTATGACT 700
```

```
217 etYrGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyrP 234
|||||
701 TTATGGCTGGCTCCGGTGGGATGATCTACTGAGACTGTGTGATCATATC 750
|||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
751 CAITGACTGTAAACAACTAAAGGTGGAAGTGTACCTTTTGGCCCGCAGT 800
|||||
251 lIleLeuIleThrSerAsnGlnThrProLeuGluThrTrpTyrSerSer 267
|||||
801 ATCTGATTACCAAGCAATCAACCCCGTTGGAATGGTACTCCTCAACTGC 850
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
|||||
851 TGTCCAGCTGTAGAAGCTCTCTATCGAGGATTACTTCTTGGTATTT 900
|||||
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGAATGTTACAGAACAACTCCACGGAGGAAGGGCCAGTTTGTCAAC 950
|||||
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCATGCTCCTGAATTTCCATATGAATTAATTAC 992
|||||
seq_name: gb_vi:AF086835
seq_documentation_block: 1768 bp DNA circular VRL 29-SEP-1998
LOCUS AF086835
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION AF086835
NID g3661518
VERSION AF086835.1 GI:3661518
KEYWORDS
SOURCE
  ORGANISM
    porcine circovirus.
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
  1 (bases 1 to 1768)
    Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
    Emergence of a new porcine circovirus
    Unpublished
  2 (bases 1 to 1768)
    Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
    Direct Submission
    Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
    Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
  source
    1..1768
      /organism="porcine circovirus"
      /strain="9741"
      /db_xref="taxon:46221"
      /codon_start=1
      /product="p35.8"
      /protein_id="AAC61739.1"
      /db_xref="PID:g3661519"
      /db_xref="GI:3661519"
      /translation="MPSKKNRSGPOPKRHWFTLNPNSENKRKKIRELPISLDFDI
      VEEEGEERTPHLQGFANFKQTFNKVLYLGARHIEKAGTDQDNKQKCYCKEKN
      LLIECGAPRSQGSQSDLSSTAVSTLLESGLTLVTVAKQHPVTFKNFRGLAEKLVSGKM
      OKRDWNTVHFIVGPPGCKSKWAANFANPETTYKPKKNKWDGYHGEKVVIDDFY
      GHLPHWDLRLCDRYPLTVKTKGTVPELARSILITSNQTPLEWTSSTAVPAVEALYR
      RITSLVFNKATEQSTEGGQFVTLSPCCPEPFYEINY"
      complement(1034..1735)
      /codon_start=1
      /product="p27.9"
      /protein_id="AAC61740.1"
      /db_xref="PID:g3661520"
      /db_xref="GI:3661520"
      /translation="MTYPRRRYRRRRHRRPSHLGOILRRRPMWLVHPRHRYRWRKNGI
      ENRLSRFTGTVKATTTTTPSWADMMRFNDIDFVPPGGGTNKISIPFEYIRKVK
      VEFWPCSPITQDROGVSTAVILDDNFVKATLYIDPVNFYSSRHTIPOPFSYHSRY
      FTPKPVLDSITDYFQPNKRNQLWLRQTSGNVHVGGLTAFENSKYDQDYNIRVTM
```

BASE COUNT 464 a 360 c 480 g 464 t
ORIGIN
VOFRENLKDPPLKP*
alignment_scores:
Quality: 1701.00 Length: 314
Ratio: 5.417 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 98.726
alignment_block:
US-09-209-961-3 x AF086835 ..
Align seg 1/1 to: AF086835 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCCAAGAGAGATGGAGAGAGCGGCCCAACACACATATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
101 GGTGTTCACGCTGATATCTCTCCGAAACCAAGCGCAAGAAATACGG 150
34 luLeuProileSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
151 AGCTCCCAATCTCCCTATTGATATTATTATTGTTGGCGAGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
201 GAGGAAGACGACACACCTCACCTCCAGGGGTTCGCTAATTTTGTGA 250
67 sGlnThrPheAsnLysValLysTyrPheValGlyAlaArgCysHisIle 84
251 GCAAACTTTTAAAGTAAGTGTATTGGTCCCGCTGCCACATCG 300
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLys 100
301 AGAAGCCAAAGGACGTATCAGCAAAATATAAGATATTGCAGTAAAG 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyArgSe 117
351 GGCAACTTACTATTATGATGTGGAGCTCTCGATCTCAAGGACACGG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
401 TGACCTGTCTACTGTGTAGTACTTGTGGAGAGCGGATCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCGTGTCAACACGACACCTGTACGTTTGTCAAAAATTTCCGGGGCTG 500
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167
501 GCTGAACCTTTTGAAGTGAGCGGGAATGCAAAAGCGTGTGGAAC 550
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrp 184
551 CAATGTACACTTATTGTGGGGCCACCTGGTGGTAAAGCAAAATGGG 600
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLys 200
601 CTGCTAATTTTGAACACCGGAACCACTACTGGAACCAACCTTAAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
651 AAGTGGTGGATGTTACCATGGTGAAGAAAGTGTGTATGTAGACTT 700
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLysCysAspArgTyr 234
701 TTATGGTGGTGGTGGATGATCTACTCAGACTGTGTGATCATATC 750
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250

751 CATTGACTGTAAATAAGGTGGAACTGTACCTTTTGGCCCGCAGT 800
251 ileLeuIleThrSerAsnGlnThrProLeuGluTyrSerSerThrAl 267
801 ATTCGATTACCAGCAATCAGACCCCGTTGGATGGTACTCTCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
851 TGTCCACGCTGTAGAAGCTCTCTATCGAGGATTACTCTCTGGTATTT 900
284 rPlyAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
901 GGAAGAATGCTACAGAAATCCACGAGGAGGGGGCCAGTTTGTCAAC 950
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
951 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 992
seq_name: gb_v1:AF055391
seq_documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from USA, complete genome.
ACCESSION AF055391
NID G3598796
VERSION AF055391.1 GI:3598796
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1. .1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.999 EcoRI No.8"
/country="USA"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:G3598807"
/db_xref="GI:3598807"
/translation="MASSTPASPAPSDILSRPQSERPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:G3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVALTVYVKVRRRLKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"
/db_xref="PID:G3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSSNATPPWIRHS"
complement(311..439)
CDS
CDS
CDS

```
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGSSASSILLAGHVAEEVLPRCCRSALVILTAHFFRFOL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKKNGRSGPQPKRWVFTLNNPSEDERKKIRELPISLFYFI
VGESENGRPHLOGFANFYKOTFNKWTGLGARCHEKAKGDOONKEVCYKREGN
LLISCGAPGQSRDLSTAVSTLLESGLVTAQHPVTFVRNPRGLAELLKYSKRM
QKRWKTNVHVIVPGGCKSKAANFADPETTYWKPFRNKWMDGIHGEVVVIDDFY
GWLWDLLRLCDRYPLVETGKGTVPELARSILITSNOTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEOSTEGGQFVTLSPCPPEFFYEINT"
674..679
complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MVTIPPLVSRWFPVCGFRVCKLISPPFAFTTPRPHNDVYIGLPI
TLHFFAHQFQSPAEISDKRYRVLCCNGHOTPALQOQTHSSROVTPLSRSTF
NK"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCILVFQSRFCIFPLTFKSSAPRKFLTNVTGCCSATVTRPL
SNKVLTAVDRLRCP"
complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDHPTAASHKSHQ"
1330..1335
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKTKQ"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIHLGFKWGVFKIFSELYIHGYTDIVLVVYTVFERSAEA
YVYHISRLG"
complement(1369..1374)
complement(join(1381..1768,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35299.1"
/db_xref="PID:g3598798"
/db_xref="GI:3598798"
```

```
/translation="MTYPRRRYRRRRHPRSHLQILRRPWLHPRHRYRWRKNGI
FNRLSRFGYTVKATVTTPSWAVMMFNIDDFVPGGGTNRKISIPFEYIRKRVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTVDYVNTSSRHTIPQPFYHSRY
FTPKPVLSDIDYFQPNKRNQNLWLRLQTSRNVHVLGTAFENSIDQDYNIRVTM
VOFRENLKDPPLKP"
BASE COUNT 453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:
  Quality: 1682.00      Length: 314
  Ratio: 5.391         Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF055391
Align seg 1/1 to: AF055391 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
398 ATGCCAGCAAGAAAGATGGAGAGCGGACCCCAACCAACATATAAGGTG 447
|||||

17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLeArgG 34
|||||
448 GGTGTTCCAGCGTGAATAATCCTCCGAAGACGAGCGCAAGAAATACGGG 497
|||||

34 LuLeuProLisLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||
498 AGCTCCCAATCTCCTATTGATTATTATTGTTGGCGAGGAGGTAAT 547
|||||

51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
548 GAGGAAGACGACACACCTCACCTCCAGGGTTCGCTAATTTTGTGAAGA 597
|||||

67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84
|||||
598 GCAAACTTTTAATAAGTGAAGTGTATTTGGTGGCGCTGCCACATCG 647
|||||

84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
648 AGAAAGCCAAAGGAACATGATCAGCAGATAAAGAAATATTGCAGTAAAG 697
|||||

101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
698 GGCACCTTACTTATTGAATGGAGCTCCTCGATCTCAAGGACACACGGAG 747
|||||

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
748 TGACCTGTCTACTGCTGTGACTACCTTGTGGAGCGGAGTCTGTGTGA 797
|||||

134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTGTGAGACGACGCCCTGTAACTGTTGTCAAGAAATTTCCGCGGGCTG 847
|||||

151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
848 GCTGAACCTTTTGAAGTGAGCGGAAAATGCAAGACGCTGATTGGAAGAC 897
|||||

167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
898 CAATGTACACGTCATTGTGGGGCCACCTGGGTGTGTTAAAGCAATACGG 947
|||||

184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCGCGGAAACCATACTGGAACACCACTAGAAAC 997
|||||

201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217
|||||
998 AAGTGGTGGGATGTTACCATGTTGGAAGAAGTGTGTTATTGATGACTT 1047
|||||

217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysCysAspArgTyrP 234
|||||
```

```

1048 TTATGGCTGGCTGGCTGGGATGATCTACTGAGACTGTGTGATCGATATC 1097
234  tLeuThrVallysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1098 CATTGACTGTAGAGACTAAGGTGAAGTGTACCTTTTGGCCGCGACT 1147
251 lleLeulleThrSerAsnGlnThrProLeuGluIutPtyrSerSerThrAl 267
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1148 ATTCTGATTACCAAGCAATCAGACCCGTTGGAATGCTACTCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1198 TGTCACAGCTGAGAGCTCTCTATCGGAGGATTACTTCTCTGGTATTTT 1247
284 tPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1248 GGAAGAAAGCTACAGAACAAATCCAGGAGAGAGGGGCGAGTCGTCACC 1297
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
1298 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 1339
seq_name: gb_vi:AF055392

seq_documentation_block:
LOCUS AF055392 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from Canada, complete genome.
ACCESSION AF055392
NID 93598808
KEYWORDS AF055392.1 GI:3598808
SOURCE
ORGANISM porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE
2 (bases 1 to 1768)
Meehan,B.M.
AUTHORS
Direct Submission
TITLE Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1..1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1010-Stoon EcoRI No.14"
/country="Canada"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35319.1"
/db_xref="PID:g3598819"
/db_xref="GI:3598819"
/translation="MASSTPASPAPSDILSSVPSQSRPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35317.1"
/db_xref="PID:g3598817"
/db_xref="GI:3598817"
/translation="MSTAQEGVTVVRLTYVPKVRRLVKMPFFLIQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35318.1"

CDS
db_xref="PID:g3598818"
/db_xref="GI:3598818"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(311..439)
/translation="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35315.1"
/db_xref="PID:g3598815"
/db_xref="GI:3598815"
/translation="MWLGSAASSILLAGHVAEEVLPRCRCRSALVILTAHFRRFOL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35309.1"
/db_xref="PID:g3598809"
/db_xref="GI:3598809"
/translation="MPSKKNRSGQPQPKRWVFTLNNPSEDEKKIRELPISLFDYFI
VGEENEGRTPHLQGSFANFVKQTFNKKVWYLGARHIEKAKGTDOONKEYCSKEGN
LLIEGAPRSQGSGLSTAVSTLLESGSLVTAEOHPVTFVRNFRGLAEILLKVGKM
OKRDKTNVHIVGPGCGSKWAANFADPTTYKPKPRNKWDGYHGEVVVYIDDFY
GWLFWDLLRLCDRYPITVETKGTGTVFPLARSILITSNQTPLEWISSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEFFEYNT"
674..679
complement(704..1018)
/translation="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35311.1"
/db_xref="PID:g3598811"
/db_xref="GI:3598811"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFTTPRWPNDVYIGLPI
TLHFPAHFQKFSQPAISIDKRYRLVLCNGHOTPALQOGTHSSROVTLPSLSRSSTF
NK"
complement(733..912)
/translation="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35312.1"
/db_xref="PID:g3598812"
/db_xref="GI:3598812"
/translation="MTCTLVQSRFCIFPLTFKSSAPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
complement(1035..1100)
/translation="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35314.1"
/db_xref="PID:g3598814"
/db_xref="GI:3598814"
/translation="MDIDHTVSVVDHPTAASHKSHQ"
1330..1335
complement(1336..1380)
/translation="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35316.1"
/db_xref="PID:g3598816"
/db_xref="GI:3598816"
/translation="MNNKNHYEVKKTKQ"
1363..1524
/translation="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35313.1"
/db_xref="PID:g3598813"
/db_xref="GI:3598813"
/translation="MVFIIHLGFKWGVKIKFSELYIHGYTDIVVLVVTVPFERSAEA
YVHISRGL"
complement(1369..1374)
/translation="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF7"
/protein_id="AAC35318.1"

polyA_site
complement(1381..1768,1..314)

```

```

/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:g3598810"
/db_xref="GI:3988810"
/translation="MTYPRRRYRRRRPRLSHGLQLRRPWLVRHRYRWRKNGI
FNRLSRTEGYTKRTVTTPGWAVDMRFKIDFVPGGNGTKISIPFYYIRKVK
VFEPWCPSPITQDGRGVSTAVILDDNFVKATALYDVPVNTSSRHTIPQPSYHSRY
FTPKPVLDSITIDYFQFNRRNQLWLRLQTSNVDHVLGAFAENSKYDQDYNIRVTMY
VQFRENLADPLKP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

```

```

alignment_scores:
  Quality: 1682.00      Length: 314
  Ratio: 5.391          Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF055392 ..
Align seg 1/1 to: AF055392 from: 1 to: 1768

```

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
398 ATGCCAGCAAGAAAGATGGAAGAGCGGACCCCAACACACATAAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLeArg 34
448 GGTGTTCCAGCGTGAATATCTTCCGAAGACGAGCGCAAGAAATACGG 497
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
498 AGCTCCCAATCTCCATATTCATATTTATTTATTTGCGGAGAGGATRA 547
51 GluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
548 GAGGAAGACCAACACCTCACCTCAGGGTTCGTAATTTTGTGAAGAA 597
67 sGlnThrPheAsnLysValLysTyrTrpLeuGlyAlaArgCysHisLe 84
598 GCAAACTTTTAAAGTGAAGTGTATTTGGTGGTGGCGGACGATCG 647
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysG 100
648 AGAAGCCAAAGGAACCTATCAGCAGATAAAGAAATATTCAGTAAGAA 697
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
698 GGCAACTTACTTATTAATGTGGAGTCTCTCGATCTCAAGCACACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
748 TGACCTGTCTACTGCTGTGAGTACCTTGTGGAGCGGAGTCTGTGTA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGly 150
798 CCGTGGCAGACGACCCCTGTACGTTTGTCAAGAAATTTCCGGGGGCTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLys 167
848 CTGAACCTTTGAAGTGAAGCGGGAATGAGAGCGGTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
898 CAATGTACACCTCATTTGGGGCCACCTGGGTGTGTTAAAGCAAAATGG 947
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
948 CTGTAATTTTTCAGACCCGGAACACATCTCGAAACACCTAGAAAC 997
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217

```

```

998 AAGTGGTGGATGTTACCATGGTCAAGAACTGCTGTATTGATGACTT 1047
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
1048 TTATGGCTGGCTCGCGGGATGATCTACTAGACTGTGTGATCATATC 1097
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
1098 CATTGACTGTAGACATAAAGGTGAACTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
1148 ATTCTGATTACCAATACAGACCCCGTTGAATGGTACTCCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgGlnIleThrSerLeuValPheT 284
1198 TGTCACGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTTGGTATTTT 1247
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
1248 GGAAGATGCTACAGAACATCCACGAGGAGGGGCCAGCTTCGTCAAC 1297
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCGCCCATGCCCTGAATTTCCATATGAAATAAATTAC 1339

seq_name: gb_v1:AF109397

seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE
  bovine circovirus.
  bovine circovirus
  Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
  1 (bases 1 to 1768)
  Hamel,A.L. and Nayar,G.P.S.
  Nucleotide sequence of a circovirus detected in cattle with various
  clinical syndromes
  Unpublished
JOURNAL
  2 (bases 1 to 1768)
  Hamel,A.L. and Nayar,G.P.S.
  Direct Submission
AUTHORS
  Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
  Laboratory, Manitoba Agriculture, 545 University Crescent,
  Winnipeg, Manitoba R3T 5S6, Canada
JOURNAL
  3 (bases 1 to 1768)
  Hamel,A.L. and Nayar,G.P.S.
  Direct Submission
AUTHORS
  Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
  Laboratory, Manitoba Agriculture, 545 University Crescent,
  Winnipeg, Manitoba R3T 5S6, Canada
REMARK
  Sequence update by submitter
COMMENT
  On Feb 2, 1999 this sequence version replaced gi:4028609.
FEATURES
  source
    1..1768
      /organism="Bovine circovirus"
      /viral
      /db_xref="taxon:85542"
      /note="BCV
      sequence was obtained from several overlapping PCRs using
      DNA extracted from various tissues from cattle; similar to
      Porcine circovirus sequence presented in GenBank Accession
      Number AF027217"
    repeat_region 13..35
      /rpt_type=tandem
      /rpt_unit=13..18
      51..995
      /note="ORF-1; similar to Porcine circovirus ORF1 encoded
      by the sequence presented in GenBank Accession Number
      CDS

```



```

AF027217.
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:g4210314"
/db_xref="GI:4210314"
/translation="MPSKNGSGPQHKRWVFTLNNPSSEDERKKIRELPISLFDYFI
VEEGNEGRTHPLOGFANFVKOTENKVKYLGARCHIEKAGTDOONKEYCSKEGN
LIECCAPGQSOGSDLSFVSTLLSESGSLVTVAEOHPVTFVFNFRGLAELLKVSQKM
QRNDKINHHVYVPPGCKSKWAANFADPTTYWPPRNKWDGTHGEVVVIDFY
GMLPWLDRLCDRLPTVETKGTGTFVFLARSILITSNQTPLEWYSSSTAVPAVEALYR
RITSLVFNKTEQSTEEGQFVTLSPPCPEFFPEYNY"
327. .332
polyA_signal
CDS
complement(357..671)
/note="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
/db_xref="PID:g4210315"
/db_xref="GI:4210315"
/translation="MTIPLVSRWFPVCGFRVCKITSSPPAFTTPRPHNDVYIGLPI
LLHFPAPHQKFSQPAEISDKRYRVLCCNGHOTPALQQGTHSSRQVTPLSLRSSSTF
NK"
CDS
complement(386..565)
/note="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AAD11931.1"
/db_xref="PID:g4210316"
/db_xref="GI:4210316"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLINVTGCCSATVTRLPL
SNKLTAVDRSLRCP"
553. .732
CDS
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:g4028610"
/db_xref="GI:4028610"
/translation="MTSLWHLGVVYKANGLLILOTRKPHGTGNEHLETSGGVMTVKKW
LILMTFMAGCRGIY"
complement(688..753)
CDS
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTVSVDHPTAASHKSHQ"
983. .988
polyA_signal
CDS
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNNKHVEYIKKTQ"
1016. .1177
CDS
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:g4028615"
/db_xref="GI:4028615"
/translation="WVFIIHLGFKWGVFKFKFSELYTHGYTDIWLWVFTVFERSAEA
YVHHSRGL"
complement(1022..1027)
polyA_signal
CDS
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:g4028612"

```

```

/db_xref="GI:4028612"
/translation="MTYPRRRRTTPRSHLGLQILRRPWLHPRHRYRWRRKNGI
FNTPLSRTFGYTKVATVTPSWADMMRFNIDDEVPDGGTNGKISIFEYRIKVK
VEFWPCSPITQDGRGVGSTAVILDNDNFVKATALTIDPYVNYSSRHTIQPFYSYHR"
complement(1522..1611)
CDS
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASSTPASPAPSDILSRLOSERPPGRWT"
1524. .1631
CDS
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="MSTAQEGVLTVALTVPYKVRERVLKMPFFLLQR"
1682. .1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:g4028617"
/db_xref="GI:4028617"
/translation="MAAGAGPSSAVTPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..12)
stem_loop
rep_origin
/note="putative; similar to the nonnucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
452 a 362 c 493 g 461 t
BASE COUNT
ORIGIN

alignment_scores:
Quality: 1682.00 Length: 314
Ratio: 5.391 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x AF109397
Align seg 1/1 to: AF109397 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCAGCAAGAAGATGGAAGACGGACCCACCAACACATATAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
101 GGTGTTACGCTGAATAATCTTCGAAGACGAGCGCAAGAAATACGG 150
|||||
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||
151 ACCTCCCAATCCCTATTATTGATTTATTTGTCGCGAGGAGGTAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAAGGACGAACACCTCACCTCCAGGGTTCGCTAATTTTGTGAAGA 250
|||||
67 sGlnThrPheAsnLysValLysTyrTrpTyrLeuGlyAlaArgCysHisIle 84
|||||
251 GCAAACTTTTAATAAGTGAAGTGGTATTGTTGGTGCCGCGCCACATCG 300
|||||
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAGCCCAAGAACTGATCAGCAGATAAGAAATATTGCAGTAAAGAA 350
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
|||||
351 GCAACTTACTTATTGAATGTGGAGCTCTCGATCTCAGGACAACGGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134

```

```
|||||
401 TGACCTGCTACTGCTGTGAGTACCTTGTGGAGCGGAGTCTGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCCTTCAGACGACGACCCCTGTACCGTTTGTCAAGAAATTCGCGGGCTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
501 GCTGACTTTTGAAGTACGCGGGAATGCAGAGCGTGTATGGAAGAC 550
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
551 CAATGTACACGCTATTGTGGGCCACCTGGTGTGTGTAAGCAATGGG 600
184 lalaAsnPheAlaAsnProGluThrThrTrpTrpLysProLysAsn 200
601 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACCACTAGAAAC 650
201 LysTrpTrpAspGlyTrpHisGlyLysValValLysValLysAspPhe 217
651 AAGTGTGGATGTTACCATGGTGAAGAGTGTGTTTATTGATGACTT 700
217 eYrGlyTrpLeuProTrpAspLeuLeuArgLysCysAspArgTrp 234
701 TTATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 750
234 rLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
751 CATGTAGTGTAGACATAAAGTGGTGGTGGTGGTGGTGGTGGTGGTGG 800
251 lIleLeuThrSerAsnGlnThrProLeuGluTrpTrpSerSerThrAl 267
801 ATTCTGATTACAGCATCAGACCCGCTTGAATGGTACTCCTCACTGC 850
267 aValProAlaValGluAlaLeuTrpArgArgIleThrSerLeuValPhe 284
851 TGTCAGAGCTGTAGAAGCTCTATATCGAGGATTACTTCCTGTTATTT 900
284 rPlysAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
901 GGAAGATGCTACAGACAAATCCAGGAGGAAGGGGCCAGTTCGTCACC 950
301 LeuSerProProCysProGluPheProTrpGluIleAsnTrp 314
951 CTTTCCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT
```

seq_name: gb_v1:AF147751

```
seq_documentation_block:
LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
VERSION 95007010
KEYWORDS AF147751.1 GI:5007010
SOURCE porcine circovirus type 2.
ORGANISM porcine circovirus type 2.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and Sorden, S.
TITLE Characterization of clinical and immune responses in young swine to experimental porcine circovirus type II infection
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yoon, K.-J. and Pogranichnyy, R.M.
TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa State University, College of Veterinary Medicine, Ames, Iowa 50011, USA
FEATURES
source 1..1768
```

```
/organism="porcine circovirus type 2"
/isolate="ISUVDL 98-15237"
/species="Sus scrofa"
/db_xref="taxon:85708"
/notes="isolated from pigs with postweaning multisystemic wasting syndrome"
51..995
/notes="ORF1: 35-36kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="pid:g5007011"
/db_xref="gi:5007011"
/translation="MPSKNGRSGPOPKRWKFTLNPSDEKRRKIRELPISLDFYFI
VGEENEGRTPHLQGFANFKVTKTKVKGWLGARCHI EKAKGTDOONKEYCSKEGN
LLIECAPRSQGSOLSTAVSLTSSGSLVTVAEQHPVTFVNRFRGLAELLKVSQGM
QKRDKNVHVIVGPGCGSKWAANFADPTTITWPKRNKWDGYHGEVVDIDFY
GWLPPDLLRLCDRQPLTVETKGTVPFLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATEQSTEEGQFVTLSPPCPEFFPEYNY"
complement(1034..1735)
/notes="ORF2: 27-28kd"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="pid:g5007012"
/db_xref="gi:5007012"
/translation="MTYPRRYRRRRRPSHLGQILRRPWLVPVHRRYVWRKNGI
FNRLSRTGYTKATVTPSPWVDMRNIIDDFVPPGGGTTNKISIPFYYRIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVTKATATIDPYNVNSRHTIPQPFSCHSRY
FTPKPVLSDTIDYFQPNKRNQLWLRLQTSRNVHDVHGLGTAFAENSIYDQDNIIRVTY
VOFRENLKDPPLKP"
BASE COUNT 450 a 362 c 494 g 462 t
ORIGIN
```

alignment_scores:

```
Quality: 1682.00 Length: 314
Ratio: 5.391 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.452
```

alignment_block:

US-09-209-961-3 x AF147751

Align seg 1/1 to: AF147751 from: 1 to: 1768

```
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCAGCAAGAAGATGGAAGAGCGGCCCAACCAACCAATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
101 GGTGTTACGCTGAATAATCTTCGGAAGACGAGCGCAAGAAATACGGG 150
34 luLeuProIleSerLeuPheAspTrpPheIleValGlyGluGlyAsn 50
151 AGTCCCAATCTCCCTATTGATTATTTATTGTTGGCAGGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
201 GAGGAAGACGACCAACCTCCACCTCCAGGGTTCGCTAATTTTGTGAAG 250
67 sClnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisIle 84
251 GCAAACTTTTAAAGTGAAGTGTATTGGGTGCCGCTCCACATCG 300
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTrpCysSerLysGlu 100
301 AGAAGCTAAAGGACTCATCAGCAGATAAAGATAATATGCACTAAGAA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyLys 117
351 GGCAACTTACTTATTGAATGTGGAGCTCTCTCGATCTCAAGGACACGGAG 400
```

```

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTGTCTACTGCTGTGAGTACCTTGTGGAGAGCGGGAGTCTGGTGA 450
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTTGCAGAGCAGCACCCTGTACGTTTGTACAGAAATTCGCGGGGCTG 500
|||||
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167
|||||
501 GCTGAACCTTTGAAGTGAGCGGGAATGCAGAGCGTGATGGAAAC 550
|||||
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTyrP 184
|||||
551 CAATGTACACGTCTATGTGGGCCACCTGGGTGGTAAAGCAATATGGG 600
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAAATTTTGCAGACCGGAACACATCTGGAACACCACTAGAAAC 650
|||||
201 LysTrpTyrAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
|||||
651 AAGTGTGGGATGTTACCATGGTGAAGAGTGGTGTATTATGATGACTT 700
|||||
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGTGGCGTGGGATGATCTACTGAGACTGTGTGATCATATC 750
|||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
751 CATTGACTGTAGAGACTAAAGGTGGAAGTGTACCTTTTGGCCGCAGT 800
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCTGATTACCAAGATACAGCCCGTGGAAAGTGTACTCCTCAACTGC 850
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCCAGCTCTAGAAGCTCTCTATCGGAGGATTAATTCCTTGGTATTT 900
|||||
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGATGCTACAGAAATCCACGGAGGAGGGGCCAGTTCGTACC 950
|||||
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCCTGACCTGAAATTTCCATATGAAATAAATAC 992
|||||

```

seq_name: gb_vi:PCAJ3185

```

seq_documentation_block: 1768 bp DNA circular VRL 03-JUL-1998
LOCUS PCAJ3185
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION AJ223185
NID g3293040
VERSION AJ223185.1 GI:3293040
KEYWORDS complete genome.
SOURCE porcine circovirus.
ORGANISM porcine circovirus
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Morozov,I.
TITLE Direct Submission
JOURNAL Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
50011, U.S.A
REFERENCE 2 (bases 1 to 1768)
AUTHORS Morozov,I., Shiranumit,T., Sorden,S.D., Halbur,P.G., Morgan,M.K.,
Yoon,K.J. and Paul,P.S.
TITLE Detection of a novel strain of Porcine circovirus in pigs with
Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL Unpublished
FEATURES Location/Qualifiers

```

```

source 1. 1768
/organism="porcine circovirus"
/virion
/isolate="ISU-31"
/species="Sus scrofa"
/db_xref="taxon:46221"
51. .995
/notes="ORF1"
/codon_start=1
/protein_id="CA11157.1"
/db_xref="PID:el310034"
/db_xref="PID:g3293041"
/db_xref="GI:3293041"
/translation="MPSKNGRSGPQPKRWFTLNNPSEDERKKIRLPLISLFDYFI
VGEENEGRTPLHGFQNFVKOTFNKVLGARCHIEKAKGTDOONKEVCSKEGN
LLIECGAPRSGQORSDLSTAVSLLESGLVTVAEQHPVTFVFRNFRGLAELLKYSKGM
QKRMDKRNHVIVGPPCGSKSKAANFADPETTYQHPKRNKWDGIHGEEVVVDDFY
GWLPMDDLLRCDRYPLTVETKGTGTVPLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKATQSTEEGGQFVTLSPCPPEFYINY"
BASE COUNT 452 a 361 c 492 g 463 t
ORIGIN

alignment_scores:
Quality: 1682.00 Length: 314
Ratio: 5.391 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.452

alignment_block:
US-09-209-961-3 x PCAJ3185
Align seg 1/1 to: PCAJ3185 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCAGCAGAGAGATGGAAGAGCGGACCCCAACACATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
101 GTGTTCACGCTGAATAATCTTCGAGACGAGCGCAAGAAATACGGG 150
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
151 AGTCCCAATCTCCCTATTGATTATTTATTGTTGGCGAGGAGGTAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
201 GAGGAAGGACGAACACCTCACCTCCAGGGGTTCGCTAATTTGTGAAG 250
67 sGlnThrPheAsnLysValLysTyrTrpLysLeuGlyAlaArgCysHis 84
251 GCAAACTTTTATAAGTGAAGTGTATTGTTGGTCCCGCTGCCACATCG 300
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
301 AGAAAGCCCAAGAACTGATCAGCAGATATAAGATATTGTCAGTAGAG 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
351 GGCAACTTACTTATTGATGTGGAGCTCCTCGATCTCAAGGACAACGGAG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
401 TGACCTGTCTACTGCTGTGAGTACCTTGTGGAGAGCGGGAGTCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCGTTGCAGAGCAGCACCCTGTACGTTTGTACAGAAATTCGCGGGGCTA 500
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrLysTh 167
501 GCTGAACCTTTGAAGTGAGCGGGAATGCAGAGCGTGATGGAAAC 550
501 GCTGAACCTTTGAAGTGAGCGGGAATGCAGAGCGTGATGGAAAC 550

```

```

167 rAsnValHisPheIleValGlyProProGlyCysGlySerLysTrpA 184
|||||
551 CAATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAAAATGGG 600
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAAATTTTGCAGACCGGAACCATACACTGGAACCACTAGAAAC 650
|||||
201 LysTrpTrpAspGlyTyrHisGlyLysValValValValIleAspAspPh 217
|||||
651 AAGTGGTGGGATGTTACCAAGTGAAGAGTGGTGTATTGATGACTT 700
|||||
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGCTGGCGGTGGATGATCTACTGAGACTGTGTGATCGATAC 750
|||||
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
751 CATTGACTGTAGACTAAAGGTGAACCTGTACCTTTTGGCCCGCAGT 800
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCTGATTACCAACATACAGCCCGTGGAAATGGTACTCTCTCACTGC 850
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCCAGCTGTAGAAGCTCTCTATCGAGGATTACTTCTCTGTGTAATTT 900
|||||
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGAATGCTACAGACAATCCACGAGGAAGGGGGCCAGTTCGTCAAC 950
|||||
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCCTGCTGCTGATTTCCATATGAAATAAATTAC 992

```

seq_name: gb_vi:AF166528

```

seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID 95739338
VERSION AF166528.1 GI:57393338
KEYWORDS Porcine circovirus.
SOURCE Porcine circovirus.
ORGANISM Porcine circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-yi Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-yi City 600, Taiwan
FEATURES
source
1. .1768
/organism="porcine circovirus"
/virion
/strain="Tainan"
/db_xref="taxon:46221"
/country="Taiwan"
51. .995
/codon_start=1
/product="putative Rep protein"
/protein_id="AAD50432.1"
/db_xref="PID:G5739339"
/db_xref="GI:5739339"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI"

```

```

VEEGNEERTPHLQGFANFVKQTFNKVWFGARCHIEKAKGTDOQNKCYCKEGN
LLIECGAPRSQSDLSSTAVSLLESGLTVABQHPVTYFVNFRLAEALLKVSCKM
OKRDKNTNVHVIVGPGGCKSKAANFADPTTYNKPNNKWDGYPHGEVVDYDFY
GWLWDLLRLCDRPLVETGKGTVPFLARSLITISNOTPLEWISSTAVPAVEALYR
RTISLVFNKNATEQSTEEGGQFVTLSPPCPEFFYEINY*
BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN
alignment_scores:
Quality: 1678.00 Length: 314
Ratio: 5.378 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.134
alignment_block:
US-09-209-961-3 x AF166528
Align seg 1/1 to: AF166528 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCACGACAGAAATGAGAGCGGACCCACCAACCAATAAAGGTG 100
|||||
17 pValPheThrLeuAsnAsnProSerGluAspLysLysLysLysLysLys 34
|||||
101 GGTGTTCACGCTCAATAATCTTCCGAAGACGAGCGCAAGAAATACGGG 150
|||||
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||
151 AGCTCCCAATCTCCCTATTGATTATTATTGTCGAGGAGGGAAT 200
|||||
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAGCGCGAACACCCACCTACAGGGTTCGCTAATTTTGTGAAGAA 250
|||||
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisLileG 84
|||||
251 GCAAACTTTTAATAAAGTGAAGTGTATTTTGTGCGCGCTGCCACATCG 300
|||||
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCGAAAGSAAACAGATCAGCAGATAAAGAAATACTGCAGTAAAGAA 350
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GGCAACTTACTATTCGAGTGTGGAGCTCTAGATCTCAAGGACACACCGAG 400
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTGTCTACTGCTGTGAGTACTCTTGTGGAGCGGGAGTCTGTGTGA 450
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTGTGAGAGACACCCCTGTAACTGTTCAGAAATTTCCCGCGGGCTG 500
|||||
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTGAATTTTGAAGTGAAGCGGGAATGCAAGCGGTGATTTGGAGAC 550
|||||
167 rAsnValHisPheIleValGlyProProGlyCysGlySerLysTrpA 184
|||||
551 CAATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAAAATGGG 600
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAAATTTTGCAGACCGGAACCATACACTGGAACCACTAGAAAC 650
|||||
201 LysTrpTrpAspGlyTyrHisGlyLysValValValValIleAspAspPh 217
|||||
651 AAGTGGTGGGATGTTACCAAGTGAAGAGTGGTGTATTGATGACTT 700
|||||
217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrP 234
|||||

```

```

701 TTATGTTGGCTGCCGTGGGATGATCTACTGAGACTGTGTGATCGATATC 750
234 rLeuThrValIysGlyGlyThrValProPheLeuAlaArgSer 250
|||||:::|||||
751 CATTCAGCTAGAGACTAAGGTGAAGTGTACCTTTTGGCCCGCAGT 800
251 ileLeuIleThrSerAsnGlnThrProLeuGluIleTyrSerThrAl 267
|||||:::|||||
801 ATTCTGATTACCAAGCAATCAGACCCGTTGGAATGGTACTCCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
|||||:::|||||
851 TGTCACAGCTAGAGAGCTCTATCGGAGGATTACTCTCTGGTATTT 900
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||:::|||||
901 GGAAGAATGCTACAGAACAATCCAGGAGGAGGGCCAGTCTCGTACC 950
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||:::|||||
951 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 992

seq_name: gb_vi:AF109399

seq documentation block:
LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE
ORGANISM porcine circovirus type 2-E.
Virus; ssDNA viruses; Circoviridae; Circovirus; #69#line
Circovirus type 2.
1 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
Unpublished
2 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Direct Submission
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
Location/Qualifiers
1. .1768
/organism="porcine circovirus type 2-E"
/virion
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
rpt_type=tandem
rpt_unit=13..18
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"

repeat_region 13..35
rpt_type=tandem
rpt_unit=13..18
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"

polyA_signal
CDS
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MTIPPLVSRWFVPCGFRVCKISSPFAFTTTPRWPHNDVY IGLPI
TLHFPAHFQFQSQAETSDKRYRVLLNGHTPALQOQTHSSRQVTPLSRSSTL
NK"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"
/translation="MTCTLVFQSRFCIFPLTFKSSAPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLSRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MYTSLWGLGVVKGANGLLILQTRKPHGTGNHLETSGGVMYTVKKW
LLMTFMAGCRGMIY"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHTVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVKKTO"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MVFILHGLFKWGFKIFSELYIHGYTDIVLVVFTVPEASAFA
YVHISGLSEHPQLIPFVILEVINSIKRFGCEVPGVVGEGGLGCMAGGVYVGV
IGLGGGLYKVVV"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MYTPRRFRRRRPRSHLQGLRRRPLVHPHRYRWKKNKI
FNARLSRTFGYTVATPSTPSWADMLRFLNDDFVPPGGGNTKISIFEYIRIKVK
VEFWPCSPITQDGRGVSSAILDDNFVIKATQTDYDYNVYSSRHTIPQPSYHSRY
FTPKPVLDSITDYFQPNKRNQLNQLSRNVDHVLGLTAFENSKYDQDYNIRVTMY
VQRFENLKDPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGLTVVTPVKYRERALKMPFLFOR"
complement(1528..1611)
/codon_start=1

```

```

/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRLPQSRPPGR"
1682..1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGAVSSSAETPPWIRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/stem_loop
rep_origin
/note="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 448 a 362 c 495 g 463 t
ORIGIN

alignment_scores:
    Quality: 1677.00      Length: 314
    Ratio: 5.392          Gaps: 0
    Percent similarity: 99.045      Percent Identity: 97.134

alignment_block:
US-09-209-961-3 x AF109399 ..
Align seg 1/1 to: AF109399 from: 1 to: 1768
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
51 ATGCCAGCAGAGAGATGAGAGAGCGGCCCAACACACATATAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
101 GGTGTTCCAGCTGAATAATCTCCGAAGACGAGCGCAAGAAATACGGG 150
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
151 AGCTTCCAATCTCCCTGTTGATTATTATTGTTGGCAGAGGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
201 GAGGAAGACGACACACCTCCACCGGGTTCGCTAATTTGTGAAAAA 250
67 sGlnThrPheAsnLysValLysTyrPheLeuGlyAlaArgCysHisIleG 84
251 GCAAACTTTTAATAAAGTGAAGTGTATCTTTGGTCCCGCTGCCACATCG 300
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
301 AGAAAGCCAAAGGAACATGATCAGCAGAGATTAAGAAATATTGCAGTAAGAA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
351 GGCACACTACTATTAGTGTGGAGCTCTCGATCTCAAGACACACGGAG 400
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
401 TGACCTGTCTACTCTGTGAGTACTCTGTTGGAGCGGGAGTCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
451 CCGTGTGACAGACACACCTGTAACTGTTGTCAGAAATTTCCCGGGGCTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
501 GCTGAACCTTTTGAAGTGACGGGAAAAATGCAGAACGCTGATTGGAAGAC 550
167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
|||||

```

```

551 CAATGTACACGTCATTGTGGGCCACCTGGGTGGTAAAGCAAAATGGG 600
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAATTTTGCAGACCCGGAACCATATCTGGAACACCACTAGAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAspPh 217
651 AAGTGGTGGGATGTTACCATGGTGAAGAACTGGTTGTTATTGATGACTT 700
217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
701 TTATGGCTGGCTCGCGTGGATGATCTACTGAGACTGTGTGATCGATATC 750
234 rLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
751 CATTGACTGTAGAGCTAAAGTGAAGCTGTACCTTTTGGCCCGCAGT 800
251 IleLeuIleThrSerAsnGlnThrProLeuGluTyrTrpTyrSerSerThrAl 267
801 ATCTGATTACCCAGCAATCACACCCCGTTGGAATGGTACCCCTCAACTGC 850
267 aValProAlaValGluAlaLeuTyrArgIleThrSerLeuValPheT 284
851 TGTCCCAAGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTCTGGTATTTT 900
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
901 GGAAGATGCTACAGAACATCCACGGAGGAGGGCCAGTTCGTCACC 950
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
951 CTTTCCCCCATGCCCTGAATTTCCATATGAATAAATTAC 992

seq_name: gb_v1:AF055393
seq_documentation_block:
LOCUS AF055393 1767 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from France, complete genome.
ACCESSION AF055393
NID 93598820
VERSION AF055393.1 GI:3598820
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus
Virus; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1767)
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1767)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1..1767
/organism="porcine circovirus"
/species="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1011 48121 EcoRI No.17"
/country="France"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35329.1"
/db_xref="PID:g3598830"
/db_xref="GI:3598830"
CDS

```

```

/translation="MASSSTPASPAPSDILSSPEQSERPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35327.1"
/db_xref="PID:g3598828"
/db_xref="GI:3598828"
/translation="MSTAQEGVLTVVRLTVYPKVRRRLVKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35328.1"
/db_xref="PID:g3598829"
/db_xref="GI:3598829"
/translation="MAAGAVSSSPVTPPWIRHI"
complement(311..439)
/notes="Predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35325.1"
/db_xref="PID:g3598826"
/db_xref="GI:3598826"
/translation="MGLGSASSILLARHVAEVLPRCCRCRSALVILTAHFFRFQI"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35320.1"
/db_xref="PID:g3598821"
/db_xref="GI:3598821"
/translation="MPSKKNRSGQPQHKRWVFTLNNPSEDERKKIRDLPLISLDYFI
VEEGNEEGRTPHLQGFANFVKQTKVYLGARCHEKAKGTQONKEYCKSGKN
LMCECAPRSQSGRDLSTAVSTLLESGSLVTVAEOHPVTFVRFGLAELLKVSCKM
QKRDWTNHYLVGPPCGSKSKWAANFADPTIYWKPPKNKWDGYHGEVYVIDFY
GNLPWDLRLCDRIYLVETVGTGTFPFLRSLIITSNQTPLEWISSTAVPAVEALYR
RITSLVFWNAEQSTEEGGQFVTLSPPCPEPFYEINY"
674..679
polyA_signal
CDS
complement(704..1018)
/notes="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35322.1"
/db_xref="PID:g3598823"
/db_xref="GI:3598823"
/translation="MVTIPLYSRWFPGFRVCKTSSPFAFTTPRPHNDVYISLPI
TLHFFPAHQKFSQPAEISDKRYRVLNCGHQTPALQQGTHSSRQVTPLSRRSSTL
HQ"
CDS
complement(733..912)
/notes="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35323.1"
/db_xref="PID:g3598824"
/db_xref="GI:3598824"
/translation="MFTCLVFQSRFCIFPLTFKSSASPRKFLTNVNGCSATVTRPLP
SNKVLTAVDRLRCP"
CDS
complement(1035..1100)
/notes="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35324.1"
/db_xref="PID:g3598825"
/db_xref="GI:3598825"
/translation="MDIDHTSVVDHPRAASHKSHQ"
1330..1335
polyA_signal
CDS
complement(1336..1380)
/notes="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35326.1"
/db_xref="PID:g3598827"
/db_xref="GI:3598827"
/translation="MNNKNHYEVKKTKQ"
complement(1369..1374)
polyA_site
CDS
complement(join(1380..1767,1..314))
/notes="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35321.1"
/db_xref="PID:g3598822"
/db_xref="GI:3598822"
/translation="MTYPRRRRRRRHRRSHLQILRRRPLVHPRRHYRWRKNGI
ENLRSLRTGCTVKTTRTPSWAVDMRMENIDFLPPGGSNRSPVFEFYIRKVK
VEFPCSPITQDGRGVSSAVIDLDFVTKATALTDPYVNYSSRHITQPFYSHSY
FTKPFNLKSTIDYFQPNKRNQLWLRQTAGNVHDVHGLTAFENSIIDQETNLRVTMT
VQPRFNLKDPPLNP"
BASE COUNT 447 a 360 c 502 g 458 t
ORIGIN
alignment_scores:
  Quality: 1676.00      Length: 314
  Ratio: 5.372          Gaps: 0
Percent Similarity: 99.363 Percent Identity: 96.815
alignment_block:
  US-09-209-961-3 x AF055393
Align seg 1/1 to: AF055393 from: 1 to: 1767
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
398 ATCCGCGAAGAAAGAAATGGAAGAGCGGCCCAACCCCAATAAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
448 GGTGTCTACTCTGATATATCTTCGAGAGCGGCGCAGAAATACGGG 497
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
498 ATCTTCCAATATCCCTATTGTATTATTTATTTGTTGGCGAGGAGGTAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
548 GAGGAAGACGACCAACACCTCACCTCCAGGGGTTCGCTAATTTTGTGAAGAA 597
67 sGlnThrPheAsnLysValLysTyrLeuGlyAlaArgCysHisIleG 84
598 GCAGACTTTTATAAAGTGAAGTGTATTTGGTGCCCTCCCATCGC 647
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
648 AGAAGCGAAGGAACAGATCAGCAGATAAAGATACTGCAGTAAAGAA 697
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArg 117
698 GGCACCTTACTGATGAGTGTGGAGCTCTAGATCTCAGGGACAACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
748 TGACCTGTCTACTGTGTGAGTACCTTGTGGAGAGCGGGAGTCTGGTGA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
798 CCGTTCGAGACGACGCCCTGTAGCTTTGTGAGAAATTCGCGGGCTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
848 GCTGAACCTTTGAAAGTGAGCGGGAATAATGCAGAACGCTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
898 TAATGTACACGCTCATTTGTGGGCCACCTGGGTGTGTTAAAGCAATGGG 947
184 LaAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
948 CTGCTAATTTTGCAGACCCGCGAACCACCATCTGGAACACCCTAGAAC 997

```


/db_xref="PID:g3598833"
/db_xref="GI:3598833"
/translation="MTYPRRRRRRRPRSHLQIILRRPWLVPRRHYRWRKNGI
FNRLSRTGTYTKRTVTPSWADMMRENINDELPGGSGNPRSVPFEYIRKVK
VEWPGSPITQDRGVSSAVILDDNFVKATALTYPYNNYSRHRITQPFVSHRY
FTKPVLDSDTDFQPNKRNQLWLRLOTTGNVDHVLGTAFENSIYDQENIRVTMY
VQPREFNKDPPLNP"
BASE COUNT 448 a 359 c 500 g 460 t
ORIGIN

alignment_scores:
Quality: 1676.00 Length: 314
Ratio: 5.372 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 96.815

alignment_block:
US-09-209-961-3 x AF055394 ..

Align seg 1/1 to: AF055394 from: 1 to: 1767

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
398 ATGCCAGCAGAGAAGATGGAAGAAGCGGACCCCAACCCATATAAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
448 GTGTCTACTCTGAATAATCCCTCCGAGACGAGCGCAAGAAATACGGG 497
34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
498 ATCTCCATATCCCTATTGATTATTTATTGTTGGCGAGGAGGTAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
548 GAGGAAGGAGCAACACCTCCACCTCAGGGGTTCTGCTAATTTGTGAAGAA 597
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84
598 GCAGACTTTTAATAAGTGAAGTGTATTGTTGGTCCCTGCCACATCG 647
84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
648 AGAAGCGAAGAGCAACAGATCAGCAGATAAAGATATCTGCAGTAAAGAA 697
101 GlyAsnLeuLeuIleGlyCysGlyAlaProArgSerGlnGlnArgSe 117
698 GCACACTTACTGATGAGTGTGAGCTCCTAGATCTCAGGACAAACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
748 TGACCTGTCTACTGTGTGAGTACCTTGTGTGAGAGCGGAGTCTGGTGA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
798 CCGTGTGCAGAGCAGACCTGTAACGTTTGTGAGAAATTTCCCGGGCGTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
848 GCTGAACTTTGAAGTGAAGCGGGAATATGCAAGACGCTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
898 TAATGTACACGTCATTGTGGGCCACCTGGGTGTGTTGTAAGCAATATGG 947
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
948 CTGCTAATTTTCAGACCCCGGAACACACATCTGGAACACCACTAGAAC 997
201 LysTTPTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
998 AAGTGGTGGGATGGTTACCATGGTGAAGAAGTGGTTGTTATTGATGACTT 1047

217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLeuLeuCysAspArgTyrP 234
1048 TTATGGCTGGCTGCCCTGGATGATCTACTGAGACTGTGTGATCGATATC 1097
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
1098 CATGTACTAGAGACTAAGGTGGAACGTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
1148 ATTCTGATTACCAGCAATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
1198 TGTCCTCAGCTGTAGAAGCTCTTTATCGGAGGATTAATCTCTGTTGTT 1247
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
1248 GGAAGAATGCTACAGAACAAATCCACGAGGAGGGGCCAGTTTCGTCA 1297
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCCCATGCTGCTGAATTTCCATATGAATAAATATAC 1339

seq_name: gb_vi:AF027217

seq_documentation_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998
DEFINITION Porcine circovirus strain pmws PCV, complete genome.
ACCESSION AF027217

NID 92689645

VERSION AF027217.1 GI:2689645

KEYWORDS

SOURCE porcine circovirus.

ORGANISM

REFERENCE 1 (bases 1 to 1768) Viruses; ssDNA viruses; Circoviridae; Circovirus.

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.

TITLE Nucleotide sequence of porcine circovirus associated with postweaning multisystemic wasting syndrome in pigs

J. Virol. 72 (6), 5262-5267 (1998)

JOURNAL 98241772

MEDLINE 2 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.S.

REFERENCE Direct Submission

TITLE Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada

JOURNAL Location/Qualifiers

FEATURES source

1..1768 /organism="porcine circovirus"

/strain="pmws PCV"

/db_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments

were sequenced; virus isolated from lung, lymph node,

spleen and tonsil tissue from pigs affected by post

weaning multisystemic wasting syndrome"

51..995

/note="ORF1; similar to Rep protein encoded by

non-pathogenic PCV, GenBank Accession Number U49186;

predicted 35.8 kDa protein"

/codon_start=1

/product="putative Rep protein"

/protein_id="AAC59462.1"

/db_xref="PID:g2689646"

/db_xref="GI:2689646"

/translation="MPSKKNGRSQPQHKRWFTLNPNSEDKKIRELPISLDFYFI
VGEENEGRTPLHQGFANFYKQTFNKRWYLGARCYIEKAGTDQNKCYCSKEGN
LLIECGAPRSQGRSDJSTAVSTLLESGSLYVAEPHTVFNRRGLAEILLKVGSKM
OKRDWNTNHYIVGPGSKSWAANFADPTTYWKPPRNKWDYHGEEVVIDFY
GWLFWDDLRLCDRYPLTVETKGTGTVPLARSILITSNQTPLEWYSSTAVPAVALYR
RITSLVFWKNKATESTEGGQFVLSPPCPPEPEINT"

117..125

misc_feature

```

polya_signal
CDS
/note="glycosylation site"
327..332
complement(357..671)
/note="ORF3; predicted 11.9 kDa protein"
/codon_start=1
/protein_id="AAC59464.1"
/db_xref="PID:g2689648"
/db_xref="GI:2689648"
/translation="MVTIPPLYSRWFPPVCGFRVCKISSPPAETTPRWPHNDVYIGLPI
TLILHPAHPQFSPAEISDKRYVLLCNGHOTPALQOQTHSSRQVTPLSLRSSTP
NK"
complement(386..565)
/note="ORF4; predicted 6.5 kDa protein"
/codon_start=1
/protein_id="AAC59465.1"
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="MTCILVFQSRFCIPPLTFKSSASPRKFLTNVTGCCAFVTRLPL
SNKVLTAVDRLSRCP"
complement(470..478)
/note="glycosylation site"
complement(688..753)
/note="ORF8; predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHTVSDHPTAAASHKSHQ"
816..824
/note="glycosylation site"
906..914
/note="glycosylation site"
983..988
complement(989..1033)
/note="ORF11; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNKNHYEVIKTKQ"
1016..1177
/note="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="MVFIFHLGKMGVFKIFSELYIHGYTDIVLVVYTFERSAEA
YVYHISRLG"
complement(1022..1027)
complement(1034..1735)
/note="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="WTYPRRYRRRRPRSHLQILRRPWLHPHRYRWRKNGI
FNRLSRFGYTKATVTPSWADMRFNIDDFVPPGGGKNKISPEYIRIRKVK
VFWRFCSPIDQDRGVSGTAVILDDNFVKATALYDYPVNYSSRHTIPQFESHSRY
FPPKVLDSIDYFQPNKRTQLWLRLQTSRNVHDVHGLGTAFENSIYDIDYINRVTMY
VQFRENLKDPPLKP"
complement(1301..1309)
/note="glycosylation site"
complement(1522..1611)
/note="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/note="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"

```

```

/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALTVYKVRERRVLMKMPFFLLQR"
1682..1741
/note="ORF7; predicted 1.9 kDa protein"
/codon_start=1
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(1732..1768.1..92))
/note="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGASILLAGHVAEVLPRCCRCRSALVILTAHFRRFOL"
Join(1746..1768.1..13)
/function="putative replication site"
Join(1762..1768.1..2)
/note="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 1676.00      Length: 314
  Ratio: 5.372          Gaps: 0
  Percent Similarity: 99.363  Percent Identity: 97.134

alignment_block:
US-09-209-961-3 x AF027217

Align seg 1/1 to: AF027217 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCGAGCAAGAAGATGGAAGAACGGACCCCAACCATATAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
101 GGTTTCACGCTGAATAATCCTCCGAGACGAGCGCAAGAAATACGGG 150
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
|||||
151 AGCTCCCAATCTCCCTATTGATTATTTATTTGTTGGCAGGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
|||||
201 GAGGAAGGAGCAACACCTCACCTCCAGGGGTTCGCTAATTTGTGAAGA 250
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIle 84
|||||
251 GCAACTTTTATAAAGTGAAGTGGTATTTGGGTGCCCGCTGCTACATCG 300
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCCAAAGGAACGTATCAGCAGAAATAAAGAAATATTCAGTAAAGA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
351 GGCAACTTACTTATTAATGATGGAGCTCCTCGATCTCAAGGACACCGAG 400
117 rAspLeuSerThrAlaValSerThrLeuGluSerGlyIleLeuValT 134
|||||
401 TGACCTGCTACTCTGTGAGTACTCTTGTGGAGCGGGAGTCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTGTGACAGACACCCCTGTACGTTTGTCAGAATTTCCGGCGGCGTG 500

```

```
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTGAACATTTTGAAGTGAAGCGGAAATGCAAGACGATGATTGGAAGAC 550
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
551 CAATGTACACGTCATGTGGGGCCACCTGGTGGTGTGTTAAAGCAATGGG 600
|||||
184 laAlaasnPheAlaasnProGluThrThrTyrTrpLysProLysasn 200
|||||
601 CTGCTAATTTTCACACCGGAACACACATCTGGAACACCTAGAAAC 650
|||||
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspaspPh 217
|||||
651 AAGTGTGGATGGTTACCATGTGTGAAGAAGTGGTGTATTGATGACATT 700
|||||
217 eYrGlyTrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGCTGGCTGGATGATCTACTGAGACTGTGTCGATCGATATC 750
|||||
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
|||||
751 CATTGACTAGAGTAAAGTGGAACTGTACCTTTTGGCCCGCAGT 800
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCGATTACCAAGCAATCAGACCGCTGGAAATGGTACTCTCAACTGC 850
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCCAGCTGTAGAAGCTCTCTATCGGAGGATTACTTCTTGGTATTTT 900
|||||
284 rPlyAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGAATGCTACAGAACAATCCACGGAGGAAGGGCCAGTTCGTCAAC 950
|||||
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 992

seq_name: gb_v1:AF117753

seq_documentation_block: 1768 bp DNA circular VRL 04-FEB-1999
LOCUS AF117753
DEFINITION Porcine circovirus type 2-D, complete genome.
ACCESSION AF117753
NID 94219093
VERSION AF117753.1 GI:4219093
KEYWORDS
SOURCE
ORIGIN
porcine circovirus type 2-D.
porcine circovirus type 2-D.
Viruses; ssDNA viruses; Circoviridae; Circovirus; #;#;ine
Circovirus type 2.
1 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Genetic Characterization of four novel type-2 Porcine circoviruses
Unpublished
REFERENCE
2 (bases 1 to 1768)
Hamel,A.L. and Nayar,G.P.S.
Direct Submission
TITLE
Submitted (05-JAN-1999) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
Location/Qualifiers
1. .1768
/organism="porcine circovirus type 2-D"
/virus
/db_xref="taxon:86385"
/note="type 2-D designation is based upon restriction
endonuclease digestion pattern
several overlapping PCR fragments were sequenced; virus
isolated from lung, mesenteric lymph node and tonsil
tissue; similar to Porcine circovirus sequence presented
in GenBank Accession Number AF027217"

FEATURES
source
1. .1768
/organism="porcine circovirus type 2-D"
/virus
/db_xref="taxon:86385"
/note="type 2-D designation is based upon restriction
endonuclease digestion pattern
several overlapping PCR fragments were sequenced; virus
isolated from lung, mesenteric lymph node and tonsil
tissue; similar to Porcine circovirus sequence presented
in GenBank Accession Number AF027217"

repeat_region
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/note="putative Rep and coat protein"
/codon_start=1
/product="ORF-1"
/protein_id="AAD12308.1"
/db_xref="PID:g4219094"
/db_xref="GI:4219094"
/translation="MPSKNGRSGPQPHKRWVTTLNPSSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKKQTFNKVYLGARCHIERKAGTDQONKEYCSKEGN
LLMECCAPRSQSDLSLSTAVSLLESGLVTAQHPVTYFVRNFRGLAELKVSCKM
OKRWKTNVHVIVGPPGCKSKAAAFADPTTYKPPRNKWDGYHGEVYVDDFY
GWLWDLLRLCDRYPLAVETKGTVPFLARSILITSNQTPLEWISSAAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPCPPEPPEYIN"
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD12309.1"
/db_xref="PID:g4219095"
/db_xref="GI:4219095"
/translation="MVTIPPLVSRWFPVCGFRVCKISSPFAFATPRWPHNDVYIRLPI
TLLHFPAPHQFSPAISDKRYRVLLCNGHOTPALKQGTGSGROVTPLSLRSSSTF
HQ"
complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD12310.1"
/db_xref="PID:g4219096"
/db_xref="GI:4219096"
/translation="MTCIFVQSRCIFPLTFKSSASPRKFLTNVTGCCSNVTRPLPL
SSKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD12311.1"
/db_xref="PID:g4219097"
/db_xref="GI:4219097"
/translation="MYTSLWGLGVAKANGLLILOTRKPHGTGNHLETSGGMVTWKKK
LLMTFMAGCRGMY"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD12312.1"
/db_xref="PID:g4219098"
/db_xref="GI:4219098"
/translation="MDTHTVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD12313.1"
/db_xref="PID:g4219099"
/db_xref="GI:4219099"
/translation="MNNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD12314.1"
/db_xref="PID:g4219100"
/db_xref="GI:4219100"
/translation="MVFIHLGFKWGFKFKFSELYIHGYTDIVLVLVVYTFERSAEA
YMWYISSSL"
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 product encoded
by the sequence presented in GenBank Accession Number
AF027217"
/codon_start=1
/product="ORF-2"
/protein_id="AAD12315.1"
polyA_signal
CDS
```

```

/db_xref="PID:94219101"
/db_xref="GI:4219101"
/translation="MTFRRRRRRRRRSHLGOILRRRPFHPRHRYRWRKNGI
ENTRSRFGYTVKATTTQPSWADLRFKIDDFVPGGNTKLSIPEYVIRKVK
VEFWPCSPITQGDVGSSAVILDNDFPKSALYIDPYVNYSSRRTIQPSYHSR"
1524..1631
/codon_start=1

```

CDS

```

/product="ORF-10"
/protein_id="AAD12317.1"
/db_xref="PID:94219103"
/db_xref="GI:4219103"
/translation="MSTAGEGVLTVVALTYPKVRERRVLKMPFFLLQR"
complement(1528..1611)
/codon_start=1

```

CDS

```

/product="ORF-6"
/protein_id="AAD12316.1"
/db_xref="PID:94219102"
/db_xref="GI:4219102"
/translation="WASSTPASPAPSDILSRLPQSERPPGR"
1682..1741
/codon_start=1

```

CDS

```

/product="ORF-7"
/protein_id="AAD12318.1"
/db_xref="PID:94219104"
/db_xref="GI:4219104"
/translation="MAAGAVSSAVTTPPWKRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)

```

stem_loop

rep_origin

```

/note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186."

```

```

BASE COUNT 452 a 358 c 497 g 461 t
ORIGIN

```

```

alignment_scores:
  Quality: 1674.00      Length: 314
  Ratio: 5.365          Gaps: 0
Percent Similarity: 99.363 Percent Identity: 96.815

alignment_block:
US-09-209-961-3 x AF117753 ..

Align seg 1/1 to: AF117753 from: 1 to: 1768

```

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
51 ATGCCCAAGAAAAGAAATGGAAGAAGCGGACCCCAACACACAAAAGGTG 100
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34
|||||
101 GGTGTTCCAGCTGAATAATCCTCCGAAGACGAGCGCAAGAAATACGGG 150
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
|||||
151 AGCTTCCAATCTCCCTTTTGTGATTATTTATTTGTTGGCAGAGGTAAT 200
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
201 GAGGAAGGACGACACCCACCCTCCAGGGTTCGTAATTTTGTGAAGAA 250
67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84
|||||
251 GCAAAACATTTAATAAGTGAATGTAATTTGGTCCCGCTGCCACATCG 300
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
301 AGAAAGCCAAAGGAACGATCAGCAGAGATAAAGAAATATTCAGTAAGAA 350
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
|||||
351 GGCACACTACTGATGGAATGTGGAGCTCCTAGATCTCAAGGACACCGGAG 400

```

```

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
|||||
401 TGACCTGTCTACCGCTGTGAGTACCTTCTTGAGAGCGGAGTCTGGTGA 450
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
451 CCGTGTGACAGACGACCCCTGTACGTTTGTGAGAAATTTCCGCGGGGTG 500
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
501 GCTGAACCTTTTGAAGTACGCGGAAAATGCAGAACGCTGATTTGAAGAC 550
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
551 GAATGTACACGTCATTGTGGGCGCACCTGGGTGTGCAAAAGCAATGGG 600
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
|||||
601 CTGCTAATTTTGCAGACCCGGAACACATCTGGAACCCACCTAGAAAC 650
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
|||||
651 AAGTGGTGGGATGTTACCATGGTGAAGAAGTGTGTTATTGATGACTT 700
217 eTyrGlyTrpLeuProTyrAspAspLeuLeuArgLeuCysAspArgTyrP 234
|||||
701 TTATGGCTGGCTCGCGTGGGATGATCTACTGAGACTGTGTGATCGGTATC 750
234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
751 CATTGACTGTAGAGACTAAAGGTGGAACGTGTACCTTTTGGCCCGCAGT 800
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
801 ATTCTGATTACCAACAATCAGACCCCGTTGGAATGGTACTCCTCGGCTGC 850
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
851 TGTCCACACTGTAGAAGCTCTCTATCGGAGGATTAATCTCTTGGTATTTT 900
284 rPlysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
|||||
901 GGAAGAATGCTACAGAACAGTCCACGAGGAGGAGGCGCAGTTCGTACC 950
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
|||||
951 CTTTCCCCCCCATGCCCTGAATTTCCATATGAATAAATAATAC 992

```



```

|||||
648 AGAAGCCAAAGGAAGTATGATCAGCAGAAATAAAGATATTCAGTAAGAA 697
|||||
101 GLyAsnLeuLeuLeuLeuGlyCysGlyAlaProArgSerGlnGlnArgSe 117
|||||
698 GGCACCTTACTTATTAAGTGGAGCTCCCGATCTCAAGGCAACCGGAG 747
|||||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||||
748 TGACCTGTCTACTGCTGTGAGTACCTGTGGAGAGCGGAGCTGTGTA 797
|||||
134 hrValAlaGluGlnHISProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTTCAGACGACGCCCTGTACGTTTTCAGAAATTCGCGGGCTG 847
|||||
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
|||||
848 GCTGAACCTTTTGAAGTGAAGCGGGAATAATCAGAAGCGTGATTGGAAGAC 897
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
898 CAATGTACAGTCACTGTGGGGCCACCTGGTGTGTAAGCAAAATGGG 947
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTrpTrpLysProProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCCCGAAACACATACACTGGAACACCTAGAAAC 997
|||||
201 LysTrpTrpAspGlyTrpHisGlyGluLysValValIleAspAspPh 217
|||||
998 AAGTGTGGGATGGTACCATGTGAAGAGTGTGTTTATTGATGACTT 1047
|||||
217 eTrpGlyTrpLeuProTrpAspLeuLeuArgLeuLysAspArgTrpP 234
|||||
1048 TTATGCTGCTGCGGGGATGATCTACTGAGACTGTGTGATGATATC 1097
|||||
234 roLeuThrValLysThrLysGlyCylThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGAGCTAAAGGTGAAGTACTACTTTTGGCCGCGAGT 1147
|||||
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTrpSerSerThrAl 267
|||||
1148 ATTCTGATTACCAAGCATCAGACCCCGTGGAAATGGTACTCCCACTGC 1197
|||||
267 aValProAlaValGluAlaLeuTrpArgArgIleThrSerLeuValPheT 284
|||||
1198 TGTCACAGCTGTAGAGCTCTCTATCGGAGGATTAATCTCTGGTATTT 1247
|||||
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
|||||
1248 GGAAGAATGCTACAGAAACAATCCACGAGGAGGCGCCAGTTCGTCAAC 1297
|||||
301 LeuSerProProCysProGluPheProTrpGluIleAsnTrp 314
|||||
1298 CTTTCCCGCCCATGCCCTGAATTTCCATATGAATAAATATAC 1339

```

seq_name: N_Geneseq_36: X35212

seq_documentation_block:

```

ID X35212 standard; DNA; 1768 BP.
AC X35212;
DE 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI; 99-246948/21.

```

```

PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 3: 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

```

alignment_scores:

```

Quality: 1682.00 Length: 314
Ratio: 5.391 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 97.452

```

alignment_block:

US-09-209-961-3 x X35212

Align seg 1/1 to: X35212 from: 1 to: 1768

```

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
398 ATGCCAGCAGAAAGATGGAAGAGCGACCCCAACCAACATATAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
448 GGTGTTCCGCTGAATATCTCTCCGAGAGAGCGCAAGAAATACGGG 497
34 luLeuProIleSerLeuPheAspTrpPheIleValGlyGluGluGlyAsn 50
498 AGCTCCCAATCTCCCTATTTATTATTATTATTGTTGGCGAGGAGGTAAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
548 GAGGAGGACCAACACCTCACCTCCAGGGGTTCGCTAATTTGTGAAGAA 597
67 sGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisIleG 84
598 GCAAACTTTTAAATAAGTGAAGTGTATTGTTGGTCCCGCTGCCACATCG 647
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTrpCysSerLysGlu 100
648 AGAAAGCCAAAGGAAGTATCAGCAGAAATAAAGATATTCAGTAAGAA 697
101 GlyAsnLeuLeuLeuGlyCysGlyAlaProArgSerGlnGlnArgSe 117
698 GGCACCTTACTTATTAAGTGGAGCTCCCGATCTCAAGGCAACCGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
748 TGACCTGTCTACTGCTGTGAGTACCTTCTGGAGAGCGGAGCTGTGTA 797
134 hrValAlaGluGlnHISProValThrPheValLysAsnPheArgGlyLeu 150
798 CCGTTCAGACGACGCCCTGTACGTTTTCAGAAATTCGCGGGCTG 847
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
848 GCTGAACCTTTTGAAGTGAAGCGGGAATAATCAGAAGCGTGATTGGAAGAC 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
948 CTGCTAATTTTGCAGACCCCGAAACACATACACTGGAACACCTAGAAAC 997
201 LysTrpTrpAspGlyTrpHisGlyGluLysValValIleAspAspPh 217
998 AAGTGTGGGATGGTACCATGTGAAGAGTGTGTTTATTGATGACTT 1047
217 eTrpGlyTrpLeuProTrpAspLeuLeuArgLeuLysAspArgTrpP 234
1048 TTATGCTGCTGCGGGGATGATCTACTGAGACTGTGTGATGATATC 1097
234 roLeuThrValLysThrLysGlyCylThrValProPheLeuAlaArgSer 250
1098 CATTGACTGTAGAGCTAAAGGTGAAGTACTACTTTTGGCCGCGAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTrpSerSerThrAl 267
1148 ATTCTGATTACCAAGCATCAGACCCCGTGGAAATGGTACTCCCACTGC 1197
267 aValProAlaValGluAlaLeuTrpArgArgIleThrSerLeuValPheT 284
1198 TGTCACAGCTGTAGAGCTCTCTATCGGAGGATTAATCTCTGGTATTT 1247
284 rPlysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
1248 GGAAGAATGCTACAGAAACAATCCACGAGGAGGCGCCAGTTCGTCAAC 1297
301 LeuSerProProCysProGluPheProTrpGluIleAsnTrp 314
1298 CTTTCCCGCCCATGCCCTGAATTTCCATATGAATAAATATAC 1339

```

184 laAlaAsnPhelAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCGGAACCATACACTGGAACACCACTAGAAC 997
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspH 217
|||||
998 AAGTGGTGGGATGTTACCATGGTGAAGAGTGGTGTATTGATGACTT 1047
217 eTyGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyrP 234
1048 TTATGGCTGGCTGGCGGGATGATCTACTGAGACTGTGTGATCATATC 1097
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGACTTAAGGTGGAAGTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
1148 ATTCGATTACCAACATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
1198 TGTCCCACTAGAACCTCTATCGGAGGATTACTTCTCTGGTATTATTT 1247
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheValThr 300
1248 GGAAGATGCTACAGAACAACTCCACGGAGGAGGGCCAGTTCTGCAC 1297
301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
1298 CTTTCCCCCATCGCCCTGAATTTCCATATGAATAAATTAC 1339

seq_name: N_Geneseq_36.X35210

seq_documentation_block:

ID X35210 standard; DNA; 1767 BP.
AC X35210;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl011-48121.
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.
PN Porcine circovirus.
PR FR2769322-AL.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Raines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14: Fig 1: 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl011-48121. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment_scores:

Quality: 1676.00 Length: 314

Ratio: 5.372 Gaps: 0
Percent Similarity: 99.363 Percent Identity: 96.815
alignment_block:
US-09-209-961-3 x X35210
Align seg 1/1 to: X35210 from: 1 to: 1767
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
|||||
398 ATGCCGAGCAGAAGAAATGGAAGACGGACCCCAACCCCATAAAGGTG 447
17 pValPheThrLeuAsnAsnProSerGluArgLysLysLysLysLysLys 34
|||||
448 GGTGTCTACTCTGTAATAATCTCCGAGACGAGCGCAGAAATACGGG 497
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50
:::|||||
498 ATCTTCCAATATCCCTATTATTGATTATTTATTGTCGAGGAGGTAAT 547
51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
|||||
548 GAGGAAGGACGAACCTCACCCTCCAGGGGTTCCGTAATTTTGTGAAG 597
67 sGlnThrPheAsnLysValLysTyrTyrLeuGlyAlaArgCysHisIleG 84
|||||
598 GCAGACTTTTAAAGTGAAGTGTATTGGTGCCCGCTGCCACATCG 647
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
648 AGAAAGCGAAAGAACAGATCAGCAGAAATAAGAACTATGCAGTAAAG 697
101 GlyAsnLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
698 GGCAACTTACTGATGAGTGTGGAGCTCTAGATCTCAGGGACAACGGAG 747
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
|||||
748 TGACCTGTCTACTGCTGTGAGTACCTTTGGAGAGCGGGAGTCTGGTGA 797
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CGTTGCCAGACGACACCCTGTAAACGTTTGTCAAGAAATTTCCGCGGCTG 847
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
848 GCTGAACCTTTGAAAGTGAGCGGGAATAATGCAGAACGCTGATTGGAAG 897
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
|||||
898 TAATGTACACGCTATTGTGGGGCCACCTGGGTGGTAAAGCAAAATGGG 947
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCGGAACCATACACTGGAACACCACTAGAAC 997
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspH 217
|||||
998 AAGTGGTGGGATGTTTACCATGGTGAAGAGTGGTGTATTGATGACTT 1047
217 eTyGlyTrpLeuProTrpAspAspLeuLeuArgLeuLysAspArgTyrP 234
1048 TTATGGCTGGCTGGCCGAGTATCTACTGAGACTGTGTGATCATATC 1097
234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGACTTAAGGTGGAAGTGTACCTTTTGGCCCGCAGT 1147
251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
|||||
1148 ATTCGATTACCAACATCAGACCCGTTGGAATGGTACTCCTCAACTGC 1197
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284


```

|||||
1198 TGCCAGCTGTAGAAGCTCTTTATCGAGGATTACTTCCTGGTATTT 1247
|||||
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
|||||
1248 GGAAGAAATGCTACAGAACAAATCCACGGAGGAGGGGCCAGTTCGTCA 1297
|||||
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
|||||
1298 CTTTCCCCCATGCGCTGAATTCATATGAATAAATTAC 1339
|||||

```

seq_name: N_Geneseq_36:X35211

seq_documentation_block:

AC X35211 standard; DNA; 1767 BP.

AD X35211.

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Impl011-48285.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

OS Porcine multisystemic wasting syndrome; pig; vaccine; ss.

PN Porcine circovirus.

PD FR2769322-AL.

PF 09-APR-1999.

PI 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI; 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PT of pregnant sows

PS Claim 14; Fig 2; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate Impl011-48285. The specification describes a preparation of type II porcine circovirus (PCV), which is particularly isolated from a lesion, from a pig with symptoms of PMWS (porcine multisystemic wasting syndrome). PCV (attenuated or inactivated), polypeptides derived from it, and vectors that express these polypeptides are all useful in

CC vaccines, suitable for administration to adult or young pigs, or to pregnant sows (for passive immunization of their offspring). DNA isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual hybridization or amplification assays. These polypeptides may also be

CC used diagnostically to detect PCV-specific antibodies, while antibodies raised against the polypeptides can be used to detect antigens, in any

CC usual immunoassay format.

CC Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

SQ

alignment_scores:

Quality	1676.00	Length	314
Ratio	5.372	Gaps	0
Percent Similarity	99.363	Percent Identity	96.815

alignment_block:

US-09-209-961-3 x X35211

Align seg 1/1 to: X35211 from: 1 to: 1767

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17

|||||

398 ATGCCAGCAAGAAGATGGAAGAGCGGACCCCAACCCCAATAAAGTG 447

|||||

17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArg 34

|||||

448 GGTGTTCACTCTGAATAATCCTCCGAAGACGAGCGCAAGAAATACGGG 497

|||||

34 lueuProIleSerLeuPheAspTyrPheIleValGlyGluGlyAsn 50

|||||

498 ATCTTCAATATCCCTATTTGATTTATTTGTTGGCGAGGAGGTAAT 547

|||||

51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67

|||||

```

548 GAGGAGGACGAAACACCTCACCTCCAGGGGTTCGCTAAATTTTGTGAAGAA 597
|||||
67 sGlnThrPheAsnLysValLysTyrThrLeuGlyAlaArgCysHisIleG 84
|||||
598 GCACACTTTTAAGTGAAGTGGTATTTGGGTGCCCGCTGCCACATCG 647
|||||
84 lUlysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
|||||
648 AGAAAGCGAAGAACAGATCAGCAGATAAAGATACTGCAGTAAGAA 697
|||||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||
698 GGCAACTTACTGAGAGTGTGGAGCTCTAGATCTCAGGCAACAACGGAG 747
|||||
117 rAspLeuSerThrAlaValSerThrLeuGluSerGlyIleLeuValT 134
|||||
748 TGACCTGTCTACTGCTGTGAGTACTTGTGGAGAGCGGAGTCTGGTGA 797
|||||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
|||||
798 CCGTTGCAGACGACGACCCCTGTAACGTTTGTCAAGAAATTCGCGGGCTG 847
|||||
151 AlaGluLeuLysValSerGlyLysMetGlnLysArgAspTyrPlystH 167
|||||
848 GCTGAACCTTTTGAAGTGAGCGGGAATGCAGAACGCTGATTGGAAGAC 897
|||||
167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTyrTPA 184
|||||
898 TAAATGTACACTTATTGTGGGCCACCTGGGTGTGTAAGAAATAATGGG 947
|||||
184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
|||||
948 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACCACTAGAAAC 997
|||||
201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
|||||
998 AAGTGGTGGGATGTTACCATGGTGAGAGTGGTGTATTATGATGACTT 1047
|||||
217 eTyrGlyTrpLeuProTyrPaspAspLeuLeuArgLysCysAspArgTyrP 234
|||||
1048 TATGGCTGGCTGCCCTGGGATGATCTACTGAGACTGTGTGATGATATC 1097
|||||
234 rOLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
|||||
1098 CATTGACTGTAGAGACTAAAGTGAAGTGTACCTTTTGGCCCGCAGT 1147
|||||
251 lIleLeuIleThrSerAsnGlnThrProLeuGluTyrTyrSerSerThrAl 267
|||||
1148 ATTCTGATTACCAGCAATCAGACCCCGTTGGAATGGTACTCTCAACTGC 1197
|||||
267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
|||||
1198 TGTCCAGCTCTAGAAGCTCTTTATCGAGGAGTACTTCCTTGGTATTTT 1247
|||||
284 rPLysAsnAlaThrLysGlnSerThrGluGluGlyGlnPheValThr 300
|||||
1248 GGAAGAAATGCTACAGAACAAATCCACGGAGGAGGGGCCAGTTCGTCA 1297
|||||
301 LeuSerProCysProGluPheProTyrGluIleAsnTyr 314
|||||
1298 CTTTCCCCCATGCGCTGAATTCATATGAATAAATTAC 1339
|||||

```

seq_name: N_Geneseq_36:X35012

seq_documentation_block:

ID X35012 standard; DNA; 1768 BP.

AC X35012.

DT 01-JUL-1999 (first entry)

DE Genomic DNA sequence of PCV strain 999PCV.

KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;

KW vaccine; ss.

OS Porcine circovirus

PN FR2769321-AL.

PD 09-APR-1999.
 PF 03-OCT-1997; 012382.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PA (YBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246947/21.
 PT New porcine circovirus from animals with porcine systemic wasting
 syndrome
 PS Claim 13; Fig 1; 35pp; French.
 CC The present sequence represents the genomic sequence of porcine
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:

Quality: 1640.00 Length: 314
 Ratio: 5.359 Gaps: 0
 Percent Similarity: 97.452 Percent Identity: 95.541

alignment_block:

US-09-209-961-3 x X35012

Align seg 1/1 to: X35012 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
 399 ATGCCCCAGCAAGAAATGGAAGAGCGGACCCCAACACACATAAAGGTG 448
 17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleArgG 34
 449 GGTGTTCACGCTGAATAATCCCTCCGAGACGCGCAAGAAATACGGG 498
 34 LuLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluArgAsn 50
 499 AGCTCCCAATCTCCCTATTATTTATTTATTTATTTATTTATTTATTT 548
 51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhValLysLy 67
 549 GAGGAANGACGAACCTTCACCTCCAGGGGTTCGCTAATTTTGTGAAGA 598
 67 sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIleG 84
 599 GCAAACTTTATTAAGTGAAGTGTATTTGGTGGCCGCTGCCACATCG 648
 84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
 649 AGAAGCCAAAGAACTGATCAGCAGATAAAGAAATATTTCAGTAAAGAA 698
 101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyClnArgSe 117
 699 GCAAACTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 748
 117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
 749 TGACCTGTCTACTGCTGTGAGTACCTTGTGGAGAGCGGGAGTCTGGTGA 798
 134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150

799 CCGTTCAGAGCAGCACCTGTAAACGTTTGTGAGAAATTTCCGCGGCTG 848
 151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
 849 GGTGAACCTTTTGAAGTGAAGCGGAAATGCAGAGCGGTGATTGGAAGAC 898
 167 rAsnValHisPheIleValGlyProGlyCysGlyLysSerLysTrpA 184
 899 CAATGTACACGCTATTGTGGGCCACCTGGGTGGTGAAGCAATATGGG 948
 184 lalaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAsn 200
 949 CTGCTAATTTTGCAGACCGGAAACACACACTACTGGAACACCACTAGA 998
 201 LysTrpTrpAspGlyTyrHisGlyGluLysValValLysValLysAsp 217
 999 AACTGTGGGATGCTTACCATGCTGAAGAGTGTGTTGTTATGATGACTT 1048
 217 eTyrGlyTyrLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
 1049 TTATGGCTGCTGCCGTGGGATGATCTACTGAGACTGTGTGATCGATATC 1098
 234 rLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
 1099 CATTGACTGTAGAGACTAAGGTGGAACCTGTACNNNNNNNGGCCCGCAGT 1148
 251 IleLeuIleThrSerAsnGlnThrProLeuGluTyrTrpTyrSerThrAl 267
 1149 ATTCTGATTACCAGCAATCAGACCCGCTTGGATGTGACTCTCTCAACTGC 1198
 267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
 1199 TGTCCAGCTGTAGAAGCTCTCTATCGAGGATTAATCTCTGTTATTTT 1248
 284 rLysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
 1249 GGAAGAATGCTACAGAACATCCACGAGAGAGGGGCCAGTTNGTCACC 1298
 301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
 1299 CTTTCCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1340

seq_name: N_Geneseq_36:X35214

seq_documentation_block:

ID X35214 standard; DNA; 1768 BP.
 AC X35214;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
 KW PCV isolate; type II porcine circovirus; PCV; PWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-AL.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI: 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows
 PS Claim 14; Fig 6; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Imp999. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral

CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:
 Quality: 1640.00 Length: 314
 Ratio: 5.359 Gaps: 0
 Percent Similarity: 97.452 Percent Identity: 95.541

alignment_block:
 US-09-209-961-3 x X35214 ..
 Align seg 1/1 to: X35214 from: 1 to: 1768

1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
 399 ATGCCACGACGAAGAAATGGAAGAGCGGACCCCAACACACATAAAGGTG 448
 17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLysArg 34
 449 GGTGTTCCAGCTGAATATCTCTCCGAAGACGAGCGCAAGAAATACGGG 498
 34 LuLeuProIleSerLeuPheAspThrPheIleValGlyGluGluGlyAsn 50
 499 AGCTCCCAATCTCCATTTGATTTATTTATTTATTTGCGGAGAGGTTWT 548
 51 GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLysLy 67
 549 GAGGAANGACGACACCTCACCTCCAGGGTTCGCTAATTTTGTGAAGAA 598
 67 sGlnThrPheAsnLysValLysTrpTrpLeuGlyAlaArgCysHisLleG 84
 599 GCRAACTTTTAATAAAGTGAAGTGTATTTGGTGGCGCGCTGCCACATCG 648
 84 LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
 649 AGAAGCCAAAGGAACGTATCAGCAGATAAAGAAATATTCAGTAAGAA 698
 101 GlyAsnLeuLeuLeuGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
 699 GGCAACTTACTATTGATGTGGAGCTCTCGATCTCAAGGACACCGAG 748
 117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyLleLeuValt 134
 749 TGACCTGTCTACTGCTGTGAGTACCTTTGGAGAGCGGAGTCTGGTGA 798
 134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
 799 CCGTGTGAGAGACACCCCTGTAACTGTTCAGAAATTTCCGGCGGGCTG 848
 151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
 849 CTTGAACTTTTGAAGTGAGCGGGAATGACAGAGCGTGTATTGGAAGAC 898
 167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrpA 184
 899 CAATGTACAGTCACTTGTGGGCGCCACCTGGGTGTGTAAGAAACAAATGG 948
 184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
 949 CTGCTAATTTTGCAGACCCGGAACCACTACTGGAACCAACCTAGAAAC 998
 201 LysTrpTrpAspGlyTyrHisGlyGluLysValValLleAspAspPh 217
 999 AAGTGGTGGGATGTTTACCATGGTGAAGAAAGTGTGTATTGATGACTT 1048
 217 eTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyrP 234
 1049 TTATGGCTGGCTGCCGGGGAATGATCTACTGAGACTGTGTGATCATCATC 1098

234 roLeuThrValLysThrLysGlyThrValProPheLeuAlaArgSer 250
 1099 CATGACTGTAGAGACTAAGGTGGAAGTGTACNNNNNNNGGCCCGCAGT 1148
 251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
 1149 ATTCTGATTACCAATCAGACCCCGTTGGAATGGTACTCTCAACTGC 1198
 267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPheT 284
 1199 TGTCCACGCTGTAGAAGCTCTCTATCGGAGGATTACTTCCTTTGGTATTT 1248
 284 rPLysAsnAlaThrLysGlnSerThrGluGlyGlyGlnPheValThr 300
 1249 GGAAGATGCTACAGAACATCCACGAGGAGGGGGCCAGTTNGTCACC 1298
 301 LeuSerProProCysProGluPheProTyrGluIleAsnTyr 314
 1299 CTTTCCCCCGCATGCCCTGAATTTCCATATGAATAAATTAC 1340

seq_name: N_Geneseq_36.X35013

seq_documentation_block:

ID X35013 standard; DNA; 1759 BP.
 AC X35013;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence which has homology to PCV sequence.
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
 KW vaccine; ss.
 OS Sus sp.
 PN PR2769321-A1.
 PD 03-APR-1999.
 PF 03-OCT-1997; 012382.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246947/21.
 PT New porcine circovirus from animals with porcine systemic wasting
 PT syndrome
 PS Disclosure: Fig 2: 35pp; French.
 CC The specification describes a genomic sequence of porcine
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format. The present sequence appears in the
 CC specification.
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment_scores:

Quality: 1447.00 Length: 315
 Ratio: 4.922 Gaps: 2
 Percent Similarity: 93.333 Percent Identity: 83.810

alignment_block:

US-09-209-961-3 x X35013/rev ..

Align seg 1/1 to reverse of: X35013 from: 1 to: 1759

```

1  MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTrp 17
   |||||
1365 ATGCCAAGCAAGAAA.....AGCGGCCGCCCAACCCCAATAAGAGGTG 1325

17  pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLeuArg 34
   |||||
1324 GGTGTTCCACCTTAATCTCTCCGAGGAGGAGAAAACAAATACGGG 1275

34  LuLeuProIleSerLeuPheAspPheIleValGlyGluGlyAsn 50
   |||||
1274 AGCTTCAATCTCCCTTTTGTATTATTGTTGCGGAGAGAAAGTTTG 1225

51  GluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPheValLys 67
   |||||
1224 GAAGAGGGTAGAACTCCTCACCCTCAGGGTTTGGGAATTTTGTAGAA 1175

67  sGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHisIle 84
   |||||
1174 GCAGACTTTTAAACAAGGTGAAGTGTATTTTGGTCCCGCTGCCACATCG 1125

84  LuLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
   |||||
1124 AGAAAGCGAAAGGAACCGACAGCAGATAAAGAACTACTGCAGTAAAGAA 1075

101 GlyAsnLeuLeuIleGlyCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||||
1074 GGCCACATATTATCGAGTGTGGAGCTCCGCGGAACAGGGGAGCGCAG 1025

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
   |||||
1024 CGACCTGTCTACTCTGTGAGTACCTTTTGGAGAGGGGTCTTTGGTGA 975

134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
   |||||
974 CTGTAGCGGAGCAGTCTCCCTGTACAGTATGTGAGAAATTTCCCGGGGCTG 925

151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
   |||||
924 GCTGAATTTTGAAGTGAGCGGAAGATGAGCAGCGTGTATTTGGAAGAC 875

167 rAsnValHisPheIleValGlyProProGlyCysGlyLysSerLysTrp 184
   |||||
874 AGCTGTACACGTATAGTGGCGCCCGCGTGTGGGAAGCCAGTGGG 825

184 laAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProProLysAsn 200
   |||||
824 CCCGTAATTTTGCTGACCTAGGGACACCTACTGGAAGCCTAGTAGAAAT 775

201 LysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAspPh 217
   |||||
774 AAGTGGTGGGATGATATCATGGAGAAGTGTGTTGTTGGATGATTT 725

217 eTyrGlyTrpLeuProTrpAspLeuLeuArgLysCysAspArgTyrP 234
   |||||
724 TTATGCTGTTACTTGGATGATCTACTGAGACTGTGTGACCGGTATC 675

234 roLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSer 250
   |||||
674 CATGACTGTAGACATAAGGGGGTACTGTCTCTTTTGGCCCGCAGT 625

251 IleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrAl 267
   |||||
624 ATTTTGTATTACAGCAATACGGCCCGCCAGGAATGTTACTCTCAACTGC 575

267 aValProAlaValGluAlaLeuTyrArgArgIleThrSerLeuValPhe 284
   |||||
574 TGTCCCACTGTAGAAGCTCTCTATCGAGGAGTACTACTTTTGAATTT 525

284 rPlyAsnAlaThrLysGlnSerThrGlu...GluGlyGlyGlnPheVal 299
   |||||
524 GGAAGACTGTGGAGACATCCAGGAGGTACCCGAGGCGCGATTGAA 475

300 ThrLeuSerProCysProGluPheProTyrGluIleAsnTyr 314

```

```

seq_name: N_Geneseq_36:tl3161
seq_documentation_block:
ID   Tl3161 standard; DNA; 1022 BP.
AC   Tl3161.
DT   23-MAY-1996 (first entry)
DE   SCSV segment 2.
KW   SCSV; promoter; transcription; transgenic plant; legume;
OS   gene expression; crop improvement; ss.
FH   Subterranean clover stunt virus isolate F.
FT   Key Location/Qualifiers
    82..924
    /*tag= a
PN   W09606932-AL.
PD   07-MAR-1996.
PF   30-AUG-1995; AU0552.
PR   30-AUG-1994; AU-007770.
PA   (CSIR ) COMMONWEALTH SCI & IND RES ORG.
PI   Boevink JS, Chu PWG, Keese PK, Khan RI, Larkin PJ;
PI   Marshall JS, Surin BP, Taylor WC, Waterhouse PM;
DR   WPI: 96-160363/16
PT   Circovirus transcription regulatory sequences and related constructs
PT   - useful in plants, esp. leguminous plants, for the modulation of
    gene expression
PS   Claim 4; Page 58; 121pp; English.
CC   The subterranean clover stunt virus (SCSV) genome has at least 7
    distinct ssDNA components, designated segments 1-7 (Tl3160-66), each
    contg. 1 major open reading frame and a non-coding region. Segment
    2 is predicted to be a viral replication-associated protein gene.
    Genetic constructs useful in the genetic engineering of plants (esp.
    legumes) comprise 1 or more heterologous gene(s) operatively linked
    to a promoter region, and in some cases also a terminator region,
    selected from segments 1-7. The transcription regulators facilitate
    expression of foreign genes in plants and also facilitate control of
    levels of gene expression in different plant tissue types.
SQ   Sequence 1022 BP; 309 A; 183 C; 245 G; 285 T;

```

```

alignment_scores:
    Quality: 176.00      Length: 319
    Ratio: 1.197        Gaps: 14
    Percent Similarity: 46.082 Percent Identity: 23.197

alignment_block:
US-09-209-961-3 x Tl3161 ..
Align seg 1/1 to: Tl3161 from: 1 to: 1022

15  LysArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLys 31
   |||||
88  AGAAGGTACTGTTTACATTAATAATACGCTACTGAGATAGAG..... 129

31  sIleArgGluLeuProIleSerLeuPhe.....AspTyrP 43
   |||||
130  ....AGAGAAACATTCCTCTCCTCTCTCAAGACGATTAACATAAT 175

43  heIleValGlyGluGlyAsnGluGlyArgThrProHisLeuGln 59
   |||||
176  TCGTTGTGCGGACGACAACTGCACTACTGGACAG...AAACACCTCCAG 222

60  GlyPheAlaAsnPheValLysLysGlnThrPheAsnLysValLysTrpTy 76
   |||||
223  GGATTTGTATCGTTCAAGAACAAAATTCGTTGTGGATTGAAGAGAA 272

76  rIeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGln 93
   |||||
273  ATTGTGTAATCGAGCTCACTGGGAAATTCGAGAGCGCGATTCTCAGA 322

```

```

93 snLysGluTyrCysSerLysGluGlyAsnLeuLeuIleGluCysGlyAla 109
|||||
323 ATCCGCGATTATTCGTGAAGAA...ACCCTAATTCCTGAATGGGATT 369
|||||
110 ProArgSerGlnGlyGlnArgSerAspLeuSerThrAlaValSerThrLe 126
|||||
370 CCGGTCAATGAAGGTTCCGAACAAGCGGAAGC... 402
|||||
126 uLeuGluSerGlyIleLeuValThrValAlaGluGlnHisProValThr 143
|||||
403 .....ATGGAGATTATGAAGAGATCCCGAAGAAA 433
|||||
143 heValLysAsnPheArgGlyLeuAlaGluLeuLysValSerGlyLys 159
|||||
434 TG.....CAATTGAAGGATCCAGATACGCTCTTCGATGTGAAGCGAAG 477
|||||
160 MetGlnLys.....ArgAs 164
|||||
478 AAATTGAAGAGGAATATTCCTGTTATGATTTTCAGAAACTCCGTC 527
|||||
164 pTTPLyThrAsnValHis..... 170
|||||
528 ATGCCAAATTGAGCTTCACAGGATTTAATGCGGGAACACAGATCGGA 577
|||||
171 .....PheIleValGlyProGlyCys...GlyLysSerLysTrp 183
|||||
578 GTATCACTCGGTCTATGTTTCAGACGAGGAGGAAGAAAGACGAGCTC 627
|||||
184 AlaAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAs 200
|||||
628 GCGAAGGAATTAATCAGG..... 645
|||||
200 nLysTrpTrpAspGlyTyrHisGlyGluLysValValValIleAspAsp 217
|||||
645 ..... 645
|||||
217 heTyrGlyTrpLeu.....ProTrpAspAspLeuLeuArg 228
|||||
646 ..TATGGATGTTTATACAGCGGAGGAGGAGCCAGGAGTATTATAT 693
|||||
229 LeuCysAspArgTyrProLeuThrValLysThrLysGlyGlyThrValPr 245
|||||
694 ATGTATGCTCAAGACCCA.....GAGAGGAATATTCGTTTGTATGTC 737
|||||
245 oPheLeuAlaArgSerIleLeuIleThrSerAsnGlnThrProLeuGluT 262
|||||
738 CAGGTGTTCTCGGAGATG.....AACTATCAGCGCATGGAGA 778
|||||
262 rp.....TyrSerSerThrAlaValProAlaValGluAla 273
|||||
779 TGTTGAAGACACAGATTTTTCAGAGTACAAATATAGCCCTGTAGATCTT 828
|||||
274 LeuTyrArgArgIleThrSerLeuValPheTrpLysAsnAlaThrLysG1 290
|||||
829 TGTATTAGGAAGTAGTTATTTAATTGTTTTCACCAACGTCGACCTGA 878
|||||
290 nSerThr 292
|||||
879 CCCCACG 885

```

seq_name: N_Geneseq_36.V71832

seq_documentation_block:

```

ID V71832 standard; DNA; 1106 BP.
AC V71832;
DT 10-FEB-1999 (first entry)
DE BBTv DNA I clone (7-4-2) nucleotide sequence.
KW Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
OS Banana bunchy top virus.
FH Key Location/Qualifiers
FT TATA_signal 1..8
FT /*tag= a

```

```

FT stem_loop 9..38
FT /*tag= b
FT CDS 62..922
FT /*tag= c
FT /product= "ORF-V2 product"
FT /transl_except= (pos:335..337, aa:Gly)
FT /transl_except= (pos:518..520, aa:Ser)
FT polyA_signal 380..385
FT /*tag= d
FT polyA_signal 901..906
FT /*tag= e

```

```

PN US5846705-A.
PD 08-DEC-1998.
PF 06-APR-1995; 418071.
PR 06-APR-1995; US-418071.
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
PI Soong T, Wu R, You L;
DR WPI; 99-059037/05.
DR P-PSDB; W87459.
PT Nucleic acids having banana bunchy top virus component sequences -
PT used to design primers for use in polymerase chain reaction
PT detection of the virus
PS Claim 1; Fig 11A-B; 27pp; English.
CC This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA I (clone 7-4-2) circular single stranded (css) DNA. The
CC invention provides nucleic acid sequences associated with BBTv that can
CC be used in a PCR technique for detecting BBTv. The nucleic acid sequences
CC (V71832 to V71833) are used as the basis for the construction of PCR
CC primers, to detect BBTv infection. The PCR technique is used for
CC detecting BBTv in plant tissues (preferably banana, especially Musa
CC species). The virus, one of the most important banana species viruses,
CC causes phloem damage and is transmitted by aphids. PCR detection gives
CC accurate, reliable and specific determination of absence or presence of
CC the virus.
SQ Sequence 1106 BP; 335 A; 225 C; 257 G; 289 T;

```

alignment_scores:

Quality: 168.50 Length: 282

Ratio: 1:170 Gaps: 13

Percent Similarity: 51.064 Percent Identity: 25.177

alignment_block:

US-09-209-961-3 x V71832

Align seg 1/1 to: V71832 from: 1 to: 1106

```

13 ProHisLysArgTyrValPheThrLeuAsnAsnProSerGluAspGluAr 29
|||||
71 CCCCTCTTAAGTGGTCTTCACCTGAATTACTCTCCGCGGCAGAGAG 120
|||||
29 gLysLysIleArgGluLeuProIleSerLeuPhe.....A 41
|||||
121 AGAAACATTT .....CTCTCTTCTGAAGGAGGAGGATGTC 158
|||||
41 spTyrPheIleValGlyCylGluGlyAsnGluGluGlyArgThrProHis 57
|||||
159 ACTACGTGTCGTCGCGACGACGAGTCCTCCGCGCACCGCCAGAACGAC 208
|||||
58 LeuGlnGlyPheAlaAsnPheValLysLysGlnThrPheAsnLysVally 74
|||||
209 CTCGAGGATATCTATCCCTGAAAAGAGATCCGCTCGCGCGATTGAA 258
|||||
74 sTrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspG 91
|||||
259 GAAGAAGTATGTTCCCGTGTCTACTGGGAGATTGCCAGAGAACGACG 308
|||||
91 lnGlnAsnLysGluTyrCysSerLysGluGlyAsnLeuLeuIleGluCys 107
|||||
309 AAGAGAATTCGAAGTACTGTTCAAAAGAA...ACCCTAATTCGGAATTA 355
|||||
108 GlyAlaProArgSerGlnGly...GlnArgSerAspLeuSerThrAlaVa 123
|||||

```

```

356 GGGTTCTCTGTTGTAATGGTCTTAATAAAGGAAATATATCGGAGATGGT 405
123 lserThrLeuLeuGlySerGlyIleLeuValThrValAlaGluGlnHisP 140
    |||:   :   :   :   :   :   :   :   :   :   :   :   :
406 TGCCTGCTCTCTGAT.....CGCATGAAATTTGACAGC 440
140 roValThrPheValLysAsnPheArgGlyLeuAlaGluLeuLysVal 156
    |||   |||   :   :   :   :   :   :   :   :   :   :
441 CTGAGATATT...CACAGATATCAATCTGTGTAATAGTTAAAAAATTC 487
157 SerGlyLysMet.....GlnLysArgAspTrpLys..... 166
    :   :   :   :   :   :   :   :   :   :   :   :
488 AAGGAGGAGTTCTTCATCTCTGCTCATAGACCATGCGCATCAATT 537
167 .....ThrAsnValHisPheIleVal 173
538 GACGGAGCAATTGACGAGGAACCGGATGATCGAAGCATCATCTGGGTCT 587
173 alGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAla 188
    |||   |||   |||   |||   |||   |||   :   :   :   :
588 ATGTCCTTATGGTAATGAGGGTAAATCAACATATGCGAAGTCACATAATC 637
189 AsnProGluThrThrTrpTrpLysProProLysAsnLys..... 201
    :   :   :   :   :   :   :   :   :   :   :   :
638 AAGAGGATTTGTTCTACACCGGGGTGGGAGGAGGAGATATCTTATT 687
202 .TrpTrpAspGlyTrpHisGlyGluLysValValIleAsp..... 215
    :   :   :   :   :   :   :   :   :   :   :   :
688 CTCCTATGTGGACGAGGATCTGACAGCATATAGTATTGATATCTCTC 737
216 .....AspPheTrpGlyTrpLeuProTrpAspLeuLeuArg 228
    |||:   |||:   |||:   |||:   |||:   |||:   :
738 GTTGTATATCAGATAT.....TTAAATTATGATGTAATAGAGCA 778
229 LeuCysAspArgTrpProLeuThrValLysThrLysGlyGlyThrValPr 245
    |||   |||   :   :   :   :   :   :   :   :   :
779 TTAAGGATAGGGTTATAGAGAGTACTAAATACAAACCCATAAAGATAGT 828
245 oPheLeuAlaArg...SerIleLeuIleThrSerAsnGlnThrPro 259
    |||:   |||:   |||:   |||:   |||:   |||:   :
829 TGAATTAGTAAATACATGTAATCGTCATGCGCAATTTCATGCT 874

seq_name: N_Geneseq_36.v71833

seq_documentation_block:
ID   V71833 standard; DNA; 1096 BP.
AC   V71833;
DT   10-FEB-1999 (first entry)
DE   BBTv DNA II clone (2-17) nucleotide sequence.
KW   Banana bunchy top virus; BBTv; PCR technique; plant; tissue: banana;
OS   Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
FH   Banana bunchy top virus.
FT   Key Location/Qualifiers
FT   TATA_signal 1..7
FT   stem_loop 8..38
FT   CDS 70..928
FT   /product= "ORF-V2 product"
FT   /transl_except= (pos:215..217, aa:Gly)
FT   polyA_signal 533..538
FT   /tag= d
FT   polyA_signal 799..804
FT   /tag= e
FT   polyA_signal 907..912
FT   /tag= f
FT   polyA_signal 1030..1035
FT   /tag= ge
PN   US5846705-A.
PD   08-DEC-1998.
PF   06-APR-1995: 418071.
PR   06-APR-1995: US-418071.

```

```

PA (BIOT-) DEV CENT BIOTECHNOLOGY.
PI Soong T, Wu R, You L;
DR WPI: 99-059037/05.
DR P-PSDB: W87460
PT Nucleic acids having banana bunchy top virus component sequences -
PT used to design primers for use in polymerase chain reaction
PS Claim 2; Columns 17-18; 27pp; English.
CC This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA II (clone 2-17) circular single stranded (css) DNA. The
CC invention provides nucleic acid sequences associated with BBTV that can
CC be used in a PCR technique for detecting BBTV. The nucleic acid sequences
CC (W71830 to W71833) are used as the basis for the construction of PCR
CC primers, to detect BBTV infection. The PCR technique is used for
CC detecting BBTV in plant tissues (preferably banana, especially Musa
CC species). The virus, one of the most important banana species viruses,
CC causes phloem damage and is transmitted by aphids. PCR detection gives
CC accurate, reliable and specific determination of absence or presence of
CC the virus.
SQ Sequence 1096 BP; 347 A; 231 C; 244 G; 274 T;

alignment_scores:
    Quality: 163.00      Length: 290
    Ratio: 1.273        Gaps: 13
    Percent Similarity: 44.138      Percent Identity: 23.793

alignment_block:
US-09-209-961-3 x V71833

Align seg 1/1 to: V71833 from: 1 to: 1096

16 ArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLys... 31
   :   :   :   :   :   :   :   :   :   :   :   :
83 AAATGGTCTTCACTCTGATTAATTCCTCCGCGAGCGAGAGAACTT 132
32 .....IleArgGluLeuProIleSerLeuPheAspTrpPheIleVal 45
   :   :   :   :   :   :   :   :   :   :   :   :
133 TCTCCTCTTCGAAGGAGGAGGATGTT.....CACTACTCTGTCTG 173
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||:   |||:   |||:   |||:   |||:   |||:   :
174 TCGGCGACGAAAGTCGCTCCGGCCACCGCCAGACAGCCTCCAGGATAT 223
62 AlaAsnPheValLysGlnThrPheAsnLysValLysTrpTrpLeuGln 78
   :   :   :   :   :   :   :   :   :   :   :   :
224 CTATCCCTGAAAAAATCAATTCGCTCGCGGATTGAAAAAGAAAGTATGG 273
78 yAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysG 95
   |||:   |||:   |||:   |||:   |||:   |||:   :
274 CTCCTGCTCTCCTGGGAGATGCGAAAGAAAGTACGACGACAGATCGCA 323
95 LuTrpCysSerLysGluGlyAsnLeuLeuIleGluCysGlyAlaProArg 111
   :   :   :   :   :   :   :   :   :   :   :   :
324 GATACTGTTGAGGAA...ACCTAGTCTTGAACCTGGTACTCCGGTG 370
112 SerGlnGlyGlnArgSerAspLeuSerThrAlaValSerThrLeuLeuGln 128
   |||:   |||:   |||:   |||:   |||:   |||:   :
371 GTTCCTGTTGGAAGG.....CGCAAGCTTCTCGA 402
128 userGlyIleLeuValThrValAlaGluGlnHisProValThrPheVal 145
403 T.....
145 ysAsnPheArgGlyLeuAlaGluLeuLysValSerGlyLysMetGln 161
   :   :   :   :   :   :   :   :   :   :   :   :
404 ..AGATTCAGAGAGAGCCCTGAGGAATTGAGATGACGATCCATCCAAAG 451
162 LysArg.....
   |||
452 TATCCGACAGATGCTTGGCAGTGAATCAATTAAGATGCCAGAAATTAATTC 501
164 .....AspTrpLysThr..... 167

```

```

502 CGAATGGTTCACGAATAAAGAAATGGCAAAATAAATTAATCAACACA 551
168 .....AsnValHisPheIleValGlyProPro 176
552 TCGAAGGTGTCTCGATGATCGAAGATATCATCTGGGTATACGGTCCCAAC 601
177 GlyCys....GlyLysSerLysTrpAlaAlaAsnPheAlaAsnProGluTh 192
602 GGAGCGCAAGGAAGTCAACCTTCGCAAGATATCATCA..... 640
192 rThrTyTrpLysProProLysAsnLysTrpTrpAspGlyTyHisGlyG 209
641 .....TTAAACCC.....GGATGGGATATATCAACGGTG 671
209 LuLys.....ValValIleAspAspPheTyGly 219
672 GAAAGACGTCGGATATGATCATCATCAATCAAGATGGATCTGTATATCAT 721
220 TrpLeuProTrpAspLeuLeuArgLeuCysAspArgTyTrpProLeuTh 236
722 TGGATT.....ATTGATATCCCAAGAGTCATTCAGATTAT..... 757
236 rValLysThrLysGlyGlyThrValProPheLeuAlaArgSerIleLeuI 253
758 .....CTGATTATGGCTTATAGAACAAATTAAGATAGAGTTTAA 800
253 leThrSerAsnGlnThrPro 259
801 TAAATCAAAATACGAAACCA 820

```

seq_name: N_Geneseq_36:V71834

```

seq_documentation_block:
ID V71834 standard; DNA; 1091 BP.
AC V71834;
DT 10-FEB-1999 (first entry)
DE BBTv DNA II clone (2) nucleotide sequence.
KW Banana bunchy top virus; BBTv; PCR technique; plant; tissue; banana;
KW Musa; phloem damage; aphid; nucleic acid detection; circular; cyclic; ss.
OS Banana bunchy top virus.
PN US5846705-A.
PD 08-DEC-1998.
PF 06-APR-1995; 418071.
PR 06-APR-1995; US-418071.
PA (BIOT-) DEV CENT BIOTECHNOLOGY.
PI Soong T, Wu R, You L;
DR WPI; 99-059037/05.
PT Nucleic acids having banana bunchy top virus component sequences -
PT used to design primers for use in polymerase chain reaction
PT detection of the virus
PS Disclosure: Fig 12A-C; 27pp; English.
CC This represents the nucleotide sequence of a banana bunchy top virus
CC (BBTV) DNA II (clone 2) and represents a subgenomic DNA. The invention
CC provides nucleic acid sequences associated with BBTv that can be used in
CC a PCR technique for detecting BBTv. The nucleic acid sequences (V71830 to
CC V71833) are used as the basis for the construction of PCR primers, to
CC detect BBTv infection. The PCR technique is used for detecting BBTv in
CC plant tissues (preferably banana, especially Musa species). The virus,
CC one of the most important banana species viruses, causes phloem damage
CC and is transmitted by aphids. PCR detection gives accurate, reliable and
CC specific determination of absence or presence of the virus.
SQ Sequence 1091 BP; 343 A; 224 C; 248 G; 276 T;

```

alignment_scores:
Quality: 156.00 Length: 135
Ratio: 2.080 Gaps: 4
Percent Similarity: 55.556 Percent Identity: 34.074

alignment_block:
US-09-209-961-3 x V71834 ..

Align seg 1/1 to: V71834 from: 1 to: 1091

```

16 ArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLys..... 30
83 AAATGGTGTCTCACTCTGAATATTCTCCGCGGAGCGAGAGACTT 132
31 .....LysIleArgGluLeuProIleSerLeuPheAspTyP 43
133 TCTCGCTCTTCTGAAGGAGGAAGATTA.....AATTACG 167
43 heIleValGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGln 59
168 CTGTCGTGCGCGACGAAGTCTCCGACGACCCGTCGGAACACCTCCAG 217
60 GlyPheAlaAsnPheValLysLysGlnThrPheAsnLysValLysTrpTy 76
218 GGATATCTATCCCTGAAGAATCTATTAAAGCTTGGTGGATTGAAGAGAG 267
76 rLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnA 93
268 GTACTCTTCGAAGGCTCACTGGGAGAGGGCGAGAGAACTGATGAACAGA 317
93 snLysGluTyrcysSerLysGlu.....Gly 101
318 ATCCGACATATCTGTTCGAAGAAACCCCTTGAAGTGGTACTCCGGTGT 367
102 AsnLeuLeuIleGluCysGlyAlaProArg.....SerG1 113
368 TCCTGGTTCGAAGAGCGCAAGCTTCGTAGATAGATCAGAGAGACCCCTG 417
113 nGlyGlnArgSerAspLeuSerThrAlaValSerThrLeuLeuGluSerG 130
418 AGGAATTGAAGATGGACGATCCATCCAAAGTATCGAGATGCTTGCACGTG 467
130 Lylle 131
468 GAATC 472

```

seq_name: N_Geneseq_36:V24093

```

seq_documentation_block:
ID V24093 standard; DNA; 1105 BP.
AC V24093;
DT 11-AUG-1998 (first entry)
DE Banana bunchy top virus component 1 DNA sequence.
KW BBTv; probe; diagnostic primer; component 1; ss.
OS Banana bunchy top virus.
PN US5756708-A.
PD 26-MAY-1998.
PF 24-FEB-1994; 202186.
PR 24-FEB-1994; US-202186.
PA (UYQU-) UNIV QUEENSLAND TECHNOLOGY.
PI Burns TM, Dale JL, Harding RM, Karan M;
DR WPI; 98-321636/28.
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
PS Disclosure: Column 29-30; 59pp; English.
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plants or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
SQ Sequence 1105 BP; 359 A; 197 C; 279 G; 270 T;

```

alignment_scores:
Quality: 156.00 Length: 238
Ratio: 1.238 Gaps: 13
Percent Similarity: 52.941 Percent Identity: 25.630

alignment_block:
US-09-209-961-3 x V24093 ..

Align seg 1/1 to: V24093 from: 1 to: 1105

```

17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAar 33
   |||:::|||||:::|||||:::|||||:::
125 TGGATGTTCCACATCAACATCCGCC..... 151

33 gGluLeuProile.....SerLeuPheAspTyrPheIleValGlyGluG 48
   ::|||:::|:::|:::|:::|:::|:::|
152 .TCACCTACCAAGTATGCGGGAAGATTCAAAATATATGTTATCAAGTGG 200

48 LuGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
   || :::::|||||:::|:::|:::|:::|:::|
201 AGAGGGACAGAGAGGGT...ACTCGTCATGTCAGGATACGTCAGATG 247

65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAar 80
   :::::|||||:::|:::|:::|:::|:::|
248 AAGAGACGAGCTCTCGAAGCAGATGAGAGCTCTTCCAGCGCGCA.. 295

80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:::|||||:::|:::|:::|:::|:::|
296 ....CACCTTGAGAAAGGAAAGGAGCCCAAGAGAGACGAGCGCATACT 341

97 ySerLysGluGlyAsnLeuLeu.....IleGluCysGlyAlaPro 110
   || |||||:::|:::|:::|:::|:::|
342 GTATGAAGGAAGACACAAAGATCGAAGTCCCTTCGAGTTTGGTGCCTTT 391

111 Arg.....SerGlnGlyGlnArgSe 117
   :::|
392 AAATGTGTCATGTAATGATATAATTTATGTCATACAGGATACGCTGA 441

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   : ::::|
442 AACGCACAACGGCCTTGGCAGTATTTATGACTGT..... 478

134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   ||| ||||| :::::
479 .....CCAAACACCTTCGATAGAGTAAGATACATTA 511

147 PheArgGlyLeuAlaGluLeuLeuLysVal.....SerGlyLy 159
   ::||| |||||:::|:::|:::|:::|
512 TACAGAGTTCACGACAGAGTTTGAATAAAACGAAGCGCATGATAGCTGAA 561

159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   | ::::|
562 AACATCCTTCAGTTCGGGACATCGGAAGTTGAAATATTATGCGGAGC 611

176 roGlyCysGlyLysSerLysTrp.....AlaAlaAsnPheAlaAsnPro 190
   || ::::|
612 CATGTCACCGAAGGATTAATTGGTCTATGCCCCAAATGGAGGAGGA 661

191 GluThrThrTyrTrpLys.....ProProLysAsnLysTrpTrpAs 204
   ::||| |||||:::|:::|:::|:::|
662 AAGACAACTTATGCAAAATATTAAATGAAGACGAAGAATGCGTTTATTC 711

204 pGly.....TyrHisGlyGluLysV 211
   : ::|||:::|
712 GCCAGGAGGAAATCAATGCGATATATGATGTTGTAATATGAGGAAA 761

211 alValValIleAsp 215
   ::|||:::|
762 TAGTTATATTGAT 775

```

seq_name: N_Geneseq_36.V24091

seq_documentation_block:

ID V24091 standard; DNA; 1103 BP.

AC V24091;

DT 11-AUG-1998 (first entry)

DE Banana bunchy top virus component 1 DNA sequence.

KW BBTV; probe; diagnostic primer; component 1; ss.

OS Banana bunchy top virus.

PN US5756708-A.

PD 26-MAY-1998.

```

PF 24-FEB-1994; 202186.
PR 24-FEB-1994; US-202186.
PA (UQO-) UNIV QUEENSLAND TECHNOLOGY.
PI Burns TM, Dale JL, Harding RM, Karan M;
DR WPI; 98-321636/28.
PT Isolated Banana bunchy top virus DNA - useful as diagnostic probes
PT and primers and for producing virus-resistant plants
PS Disclosure; Column 25-28; 59pp; English.
CC This sequence represents a DNA fragment from the component 1 of
CC the Banana bunchy top virus (BBTV). The DNA molecules can be used as
CC diagnostic probes or primers or can be inserted into plasmids or other
CC organisms e.g. to produce virus-resistant plants or to act as promoters,
CC enhancers or termination signals.
SQ Sequence 1103 BP; 355 A; 192 C; 277 G; 279 T;

```

alignment_scores:

Quality: 152.00 Length: 224
Ratio: 1.267 Gaps: 12
Percent Similarity: 53.571 Percent Identity: 25.000

alignment_block:

US-09-209-961-3 x V24091 ..

Align seg 1/1 to: V24091 from: 1 to: 1103

```

17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAar 33
   |||:::|||||:::|||||:::|||||:::
123 TGGATGTTCCACATCAACATCCCGCT..... 149

33 gGluLeuProile.....SerLeuPheAspTyrPheIleValGlyGluG 48
   ::|||:::|:::|:::|:::|:::|:::|
150 .TCGTCACCAAGTATGCGGAGTAAATATATGTTATCAAGTGG 198

48 LuGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
   || :::::|||||:::|:::|:::|:::|:::|
199 AGAGGGACAGAGGGT...ACTCGTCATGTCAGGATACGTCAGATG 245

65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAar 80
   ::||| |||||:::|:::|:::|:::|
246 AAGAGACGAGTTCCTCGAAACAGATGAGAGCTTCTTCCAGCGCGCA.. 293

80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:::|||||:::|:::|:::|:::|:::|
294 ....CACCTTGAGAAACGAAGGAGCGCCAGGAAGACGAGCGGTACT 339

97 ySerLysGluGlyAsnLeuLeu.....IleGluCysGlyAlaPro 110
   || |||||:::|:::|:::|:::|:::|
340 GTATGACGAGATACAGATCGNAGTCCCTTCGAGTTTGGTCTTTT 389

111 Arg.....SerGlnGlyGlnArgSe 117
   :::|
390 AAATGTGTCATGTAATGATAATTTATTCATGTCATACAGGATACGCTGA 439

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   : ::::|
440 AACGCATAAACGGCCTCTCGAATATTATATGACTGT..... 476

134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   ||| ||||| :::::
477 .....CCGAATACCTTCGACAGGAAGTAAGATACATTA 509

147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
   ::||| |||||:::|:::|:::|:::|
510 TACAGAGTTCACGACAGAGTTTGAATAAAACGAAGCGCATGATAGCTGAA 559

159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   | ::::|
560 GACATCCTTCAATGATGATGACATCGAATAGTAAGTAAGTAATATGCGGAGC 609

176 ro.....GlyCysGly 179
   |||
610 CATGTTATCGAAGGATTAATTTGGGTCTACGGCCCAAAATGGAGCGGAGGA 659

```



```
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   |||
   |||
200 GATAAAA.....TATATGGTAT 216

45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   ||| :|||:||||| ||| :|||:|||||:
217 ATCAAGTGGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGTTAT 263

62 AlaAsnPheValLysLysLysLysPheAsnLysValLys...TrpTyrLe 77
   :|||:|||||:|||||:|||||:|||||:
264 GTCGAGATGACAGACGAGGCTCTGAGCAGATGAGAGGCTTCTTCCC 313

77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   ||||| :|||:||||| ||||| :|||:|||||:
314 AGGCCA.....CACCTTGAGAAAGGAGGCCAAGCAAGAGGCC 357

94 ysGluTyrCysSerLysGlu..... 100
   :|||:||||| |||||
358 GGTCACTACTGTATGAAGGAGATACAAGAATCGAAGGTCCTTCGAGTT 407

101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||:|||||:|||||:|||||:|||||:
408 GGTTCATTAAATGTCATGT.....AATGA 433

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   :|||:||||| :|||:|||||:
434 TAATTTATTTGTCATACAGGATATCGTGGAACG..... 470

134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
   :|||:|||||:|||||:|||||:
471 .....CACAAAGGCTTTGGAGTATTATATGATTGCTCTAACACC 512

148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
   ||||| ||| :|||:|||||:|||||:
513 TTCGATGAAGTAAAGGATACATTATACAGATACAAAGCAGAGATGAATA 562

162 s.....ArgAspTyrLysThrAsn..... 168
   |||:|||||:|||||:|||||:
563 AACGAGCGGATGAATAGCTGGAGAACTCTTCAGTCTGCTGGACATCAG 612

169 .....ValHisPheIle 172
   :|||:|||||
613 AGGTGGAGATATCATGCGCAGCATGTCATCGGAGATATATTGGGTC 662

173 ValGlyProProGlyCys...GlyLysSerLysTyrPheAlaAsnPheAl 188
   ||||| ||| :|||:|||||:|||||:
663 TATGCCCAATGGAGGAGAGAAAGACAACGTATGCAAAACATCTAAT 712

188 aAsnProGluThrThrTyrTyrLysProLysAsnLysTyrTrpAsp. 204
   :|||:|||||:|||||:|||||:
713 GAAGACGAGAAATCGTTTTTATCT...CCAGGAGGAAATCATTTGGATA 759

205 .....GlyTyrHisGlyGlyValValValIleAspAspPheTyr 218
   ||||| ||| :|||:|||||:|||||:
760 TATGTAGACTGTATAATTACGAGGATATTGTATATTGAT..... 800

219 GlyTrpLeuPro.....TrpAspLeuLeuArgLeuCysAspArgTy 233
   :|||:||||| :|||:|||||
801 .....ATTCCAAGATGCAAGAGGATTATTA.....AATTA 832

233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
   ||| :|||:|||||:|||||:
833 TGGGTATTAGAGGAATTAAGAATGGAATAATTCAAAGCGGGAATATG 882

245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
   ||||| ||| :|||:|||||:|||||:
883 AACCGTTTTGAAGATAGATAATATGCGAAGTCATTGTAATGGCTAAC 932

257 GlnThrProLeuGlu 261
   ||| |||
933 TTCCTTCGGAAGAA 947
```

OM of: US-09-209-961-3 to: Issued_Patents_NA:* out_format : pfs

Date: Dec 31, 1999 12:26 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

-MODEL=frame1 p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09209961/runat_22121999_101043_17921/app_query.fasta.1
-DB=Issued_Patents_NA -OFMT=fastap -SUFFIX=rn1 -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.000
-GAPOP=6.000 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961
-RCPU=6 -ICPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-209-961-3

Query length: 314

Database: Issued_Patents_NA*

Database sequences: 195662

Database length: 53930012

Search time (sec): 88.010000

```

score_list:
sequence      Strd Orig      zscore      Escore Len      Documentation
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-418-071-3 + 168.50 312.27 5.4e-10 1106
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-418-071-4 + 163.00 301.20 2.2e-09 1096
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-418-071-5 + 156.00 287.03 1.4e-08 1091
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-18 + 156.00 286.87 1.4e-08 1105
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-16 + 152.00 278.76 3.9e-08 1103
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-17 + 152.00 278.75 3.9e-08 1104
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-9 + 147.50 269.53 1.3e-07 1111
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-10 + 147.50 269.53 1.3e-07 1111
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-24 + 147.50 269.53 1.3e-07 1111
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-11 + 144.50 263.44 2.8e-07 1110
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-15 + 144.50 263.43 2.8e-07 1111
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-13 + 143.50 261.42 3.6e-07 1109
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-12 + 143.50 261.40 3.7e-07 1111
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-14 + 140.50 255.31 8.0e-07 1110
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-674-852-1 + 111.00 172.77 0.0316 7400
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-416-603-1 + 105.00 154.99 0.3092 11832
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-596-864-2 + 86.00 145.28 1.07 1042
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-989-667-2 + 86.00 145.28 1.07 1042
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-852-260-1 + 95.50 141.21 1.81 7432
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-461-503-1 + 95.50 141.21 1.81 7432
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-416-523-1 + 91.00 140.34 2.02 3708
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-185-232A-1 + 90.00 138.31 2.63 3708
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-646-590B-39 + 81.50 134.61 4.22 1185
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-137-175A-4 + 81.00 127.49 10.51 1978
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-702-344-20 + 78.50 126.62 11.76 1389
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-482-142-132- + 73.00 126.49 11.96 549
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US95-11859-2 + 88.00 125.07 14.33 8010
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-633-760-47 + 80.50 124.55 15.32 2325
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-336-408B-5 + 78.50 124.48 15.46 1662
/cgn2_6/ptodata/1/ina/PCTUS9_COMB.seq:PCT-US91-00399-5 + 78.50 124.48 15.46 1662
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-019-870-2 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-019-870-7 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-019-870-10 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-633-760-43 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-633-760-45 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-633-760-49 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-633-760-51 + 79.50 122.52 19.89 2325
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-019-870-9 + 79.50 122.28 20.82 2373
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-167-854-1 + 86.50 122.16 20.82 7917
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-533-306A-5 + 80.00 121.84 21.69 2680
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-742-923A-5 + 80.00 121.84 21.69 2680
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-878-563A-2 + 78.50 121.81 21.78 2080

```

/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-140-797-1 - 75.50 121.61 22.34 1267
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-486-670A-1 - 75.50 121.61 22.34 1267
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-159-784-1 - 82.00 121.05 24.01 4031

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-418-071-3

seq_documentation_block:

: Sequence 3, Application US/08418071
: Patent No. 5846705
: GENERAL INFORMATION:
: APPLICANT: Wu, Rey-Yuh
: APPLICANT: You, Li-Ru
: APPLICANT: Soong, Tai-Seng
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
: TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS AND METHOD FOR DETE
: TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
: NUMBER OF SEQUENCES: 18
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Neave
: STREET: 1251 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: United States
: ZIP: 10020
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Word Perfect 5.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/418,071
: FILING DATE: 06-APR-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Haley Jr., James F.
: REGISTRATION NUMBER: 27,794
: REFERENCE/DOCKET NUMBER: DCB-1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 596-9000
: TELEFAX: (212) 596-9090
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1106 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: circular
: MOLECULE TYPE: genomic DNA
: DESCRIPTION: desc= "BBTV DNA I (clone 7-4-2)"
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Banana Bunchy Top Virus (BBTV)
: US-08-418-071-3

alignment_scores:
Quality: 168.50 Length: 282
Ratio: 1.170 Gaps: 13
Percent similarity: 51.064 Percent Identity: 25.177

alignment_block:
US-09-209-961-3 x US-08-418-071-3

Align seg 1/1 to: US-08-418-071-3 from: 1 to: 1106

13 ProHisLysArgTrpValPheThrLeuAsnProSerGluAspGluAr 29
71 CCCTCTTAATGTTGCTACTTCTCTCCGCGCAGAG 120
29 gLysLysIleArgGluLeuProIleSerLeuPheA 41
121 AGAAACATTTCTCTCTTCTGAAGGAGGAGGATGTC 158

```
41 spTyrPheLeValGlyGluGlyAsnGluGluGlyArgThrProHis 57
||| :|||:|||||
159 ACTACGCTGTCGTCGCGAGAGTCGCTCGGCCACCGCCAGAACAC 208
||| :|||:|||||
58 LeuGlnGlyPheAlaAsnPheValLysGlnThrPheAsnLysVally 74
||| :|||:|||||
209 CTCAGGATATATATATATATATATATATATATATATATATAT 258
||| :|||:|||||
74 sTyrLeuGlyAlaArgCysHisLeuGlyAlaLysGlyThrAspG 91
||| :|||:|||||
259 CAGACGATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 308
||| :|||:|||||
91 LngLysLysGlyTyrCysSerLysGlyAsnLeuLeuLeuGlyCys 107
||| :|||:|||||
309 AAGAGAAATTCGAAGTACTGTCAAAAGAA...ACCCTAATTCGAATTA 355
||| :|||:|||||
108 GlyAlaProArgSerGlnGly...GlnArgSerAspLeuSerThrAla 123
||| :|||:|||||
356 GGGTTTCCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 405
||| :|||:|||||
123 lSerThrLeuLeuGluSerGlyLeuValThrValAlaGluGlnHis 140
||| :|||:|||||
406 TGTCTGTTCTCTGAT.....CGCATGAAATTAAGACGC 440
||| :|||:|||||
140 roValThrPheValLysAsnPheArgGlyLeuAlaGluLeuLysVal 156
||| :|||:|||||
441 CTGAGATATT...CACAGATATCAATCTGTAATAAGTTAAAAAATTC 487
||| :|||:|||||
157 SerGlyLysMet.....GlnLysArgAspTrpLys..... 166
||| :|||:|||||
488 AAGGAGGAGTTCGTTTCATCTGCTGATAGACCATCGGAGATCAATT 537
||| :|||:|||||
167 .....ThrAsnValHisPheIleVal 173
||| :|||:|||||
538 GACGGAGGCAATTCACGAGAACCCGATGATCGAAGCATCATCTGGGTCT 587
||| :|||:|||||
173 alGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAla 188
||| :|||:|||||
588 ATGTCCTTATGTAATGAGGTAATCAACATATGCGAAGTCACTAATC 637
||| :|||:|||||
189 AsnProGluThrThrTrpLysProLysAsnLys..... 201
||| :|||:|||||
638 AAGAAGGATGTTCTACACGAGGGTGGGAAGAGAGATATCTTATT 687
||| :|||:|||||
202 .TrpTrpAspGlyTyrHisGlyGluLysValValIleAsp..... 215
||| :|||:|||||
688 CTCTATGTCGAGGAGGATCTGACAGCATATAGTATTGATATTCCTC 737
||| :|||:|||||
216 .....AspPheTyrGlyTrpLeuProTrpAspAspLeuLeuArg 228
||| :|||:|||||
738 GTTGTAATCAGGATTAT.....TTAATATGATGTAATAGAGCA 778
||| :|||:|||||
229 LeuCysAspArgTyrProLeuThrValLysThrLysGlyThrValPr 245
||| :|||:|||||
779 TTAAGGATAGGTTATAGAGAGTACTATAATACAAACCCATAAAGTAGT 828
||| :|||:|||||
245 oPheLeuAlaArg...SerIleLeuIleThrSerAsnGlnThrPro 259
||| :|||:|||||
829 TGAATTAGTGAATAATACATGTAATCGTCATGGCGAATTCATGCT 874
||| :|||:|||||
seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq;us-08-418-071-4
```

```
seq_documentation_block:
; Sequence 4, Application us/08418071
; Patent No. 5846705
```

```
; GENERAL INFORMATION:
```

```
; APPLICANT: Wu, Rev-Yuh
```

```
; APPLICANT: You, Li-Ru
```

```
; APPLICANT: Soong, Tai-Seng
```

```
; TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF TWO CIRCULAR ssDNA
```

```
; TITLE OF INVENTION: ASSOCIATED WITH BANANA BUNCHY TOP VIRUS
```

```
; TITLE OF INVENTION: BANANA BUNCHY TOP VIRUS
```

```
; NUMBER OF SEQUENCES: 18
```

```
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10020
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 5.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,071
; FILING DATE: 06-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Haley Jr., James F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: DCB-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 596-9000
; TELEFAX: (212) 596-9090
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1096 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; MOLECULE TYPE: genomic DNA
; DESCRIPTION: /desc "BBTV DNA II (clone 2-17)"
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Banana Bunchy Top Virus (BBTV)
; US-08-418-071-4
```

alignment_scores:

```
Quality: 163.00 Length: 290
Ratio: 1.273 Gaps: 13
Percent similarity: 44.138 Percent identity: 23.793
```

alignment_block:

```
US-09-209-961-3 x US-08-418-071-4
```

```
Align seg 1/1 to: US-08-418-071-4 from: 1 to: 1096
```

```
16 ArgTrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLys... 31
:||||| :|||||:|||||
83 AAATGGTGCTTCACTCTGATTTATCTCCGCGGAGCGAGAGACTT 132
||||| :|||||:|||||
32 .....IleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
:||||| :|||||:|||||
133 TCTCGCTCTTCTGAAGGAGGAGGATGTT.....CACTACTCTGTCG 173
||||| :|||||:|||||
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
||||| :|||||:|||||
174 TCGCGCAGGAAGTCGTCGCGGCCACCGCCAGAGACCTCCAGGGATAT 223
||||| :|||||:|||||
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLysTrpTyrLeuG1 78
:||||| :|||||:|||||
224 CTATCCCTGAAAAAATCAATTCGCTCGCGGATGAAAAAGAGATATGG 273
||||| :|||||:|||||
78 yAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysG 95
:||||| :|||||:|||||
274 CTCCTGCTCCTGAGGATGCGAAGGAGGAGGAGGAGGAGGAGGAGGAG 323
||||| :|||||:|||||
95 luTyrCysSerLysGluGlyAsnLeuLeuIleGluCysGlyAlaProArg 111
:||||| :|||||:|||||
324 GATACTGTGGAAGGAA...ACCCTAGTCTTGAACCTGGGTACTCCGGTG 370
||||| :|||||:|||||
112 SerGlnGlyGlnArgSerAspLeuSerThrAlaValSerThrLeuLeuG1 128
||||| :|||||:|||||
```



```

seq_documentation_block:
; Sequence 18, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1105 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-18

alignment_scores:
Quality: 156.00 Length: 238
Ratio: 1.238 Gaps: 13
Percent Similarity: 52.941 Percent Identity: 25.630

alignment_block:
US-09-209-961-3 x US-08-202-186-18 ..

Align seg 1/1 to: US-08-202-186-18 from: 1 to: 1105

17 TrpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
125 TGGATGTTCCACCATCAATCCGCC..... 151

33 gGluLeuProIle.....SerLeuPheAspTyrPheIleValGlyCluG 48
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
152 .TCACACACAGTGCAGGAGAGGTTCAATATATGTTATCAAGTGG 200

48 luGlyAsnGluGlyArgTyrProHisLeuGlnGlyPheAlaAsnPhe 64
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
201 AGAGGGACAGAGAGGGT...ACTCGTCATGTCAGGATACGTCGAGATG 247

65 ValLysLysGlnThrPheAsnLysValLys...TrpTyrLeuGlyAlaAr 80
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
248 AAGACAGACAGCTCTGAGACAGATCAGAGGCTTCTCCAGCGCA... 295

80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysLysTyrC 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
296 ....CACCTTGAGAAACGAAAGGAGGAGCAAGCAAGAGACGACGCGCAT 341

97/yysSerLysGluGlyAsnLeu.....lleGluCysGlyAlaPro 110

```

```

342 GTATGAAGGAGACACAAAGATCGAAGGTCCTTCGAGTTGGTGCCTTT 391
111 Arg.....SerGlnGlyGlnArgse 117
392 AAATTGTCATGTAATGATAATTTATTTGATCATACAGGATATGCGTGA 441
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
442 AACGCACAAACGCCCTTGGAGTATTTATATGAGTGT..... 478
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
479 .....CCAAACACCTTCGATAGTAAGTAAGGATACATTA 511
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
512 TACAGAGTTCACGACAGAGTTGAATAAAACGAGCGGATGAATAGCTGGA 561
159 sMetGlnLysArgAspTyrLysThrAsnValHisPheIleValGlyProp 176
562 AACATCCTTCAGTTCGTCGACATCGGAGTGAATAATATATGCGGAGC 611
176 roGlyCysGlyLysSerLysTrp.....AlaAlaAsnPheAlaAsnPro 190
612 CATGTCACCGAAGATAATTGGTCTCTATGCCCAATGGAGGAGGAAGGA 661
191 GluThrThrTyrTrpLys.....ProProlAsnLysTrpTrpAs 204
662 AAGACAACTTATCAAAATATTTAATGAAGACGAGAAATCGTTTATTC 711
204 pGly.....TyrHisGlyGlyLysV 211
712 GCCAGGAGGAAATCATGTCATATATGATGATATTAATATGAGGAAA 761
211 alValValIleAsp 215
762 TAGTATATTTGAT 775

seq_name: /cgn2_6/ptodata/l/ina/5B_COMB.seq:US-08-202-186-16

seq_documentation_block:
; Sequence 16, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKI
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399

```

```
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1103 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-16

alignment_scores:
  Quality: 152.00      Length: 224
  Ratio: 1.267        Gaps: 12
  Percent Similarity: 53.571  Percent Identity: 25.000

alignment_block:
US-09-209-961-3 x US-08-202-186-16 ..
Align seg 1/1 to: US-08-202-186-16 from: 1 to: 1103

17 TtpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
123 TGGATGTTCCACCATCAACATCCCGCT..... 149
33 gGluLeuProfile.....SerLeuPheAspTyrPheIleValGlyGlu 48
150 .TCGCTACCAGTGATCGGGATGAGTTAAATATATGTTATCATCAAGTGG 198
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
199 AGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGATACGTCGAGATG 245
65 ValLysLysGlnThrPheAsnLysValLys...TtpTyrLeuGlyAlaAr 80
246 AAAGACGAGAGTCTCTGAACACAGATGAGAGGCTTCTCCAGGCGCA.. 293
80 gCysHisIleGluLysAlaLysGlyThrAspGlnAsnLysGluTyrC 97
294 ...CACCTTGAGAACGAGGAGGAGCCAGGAAGACACGGGTTACT 339
97 ysSerLysGluGlyAsnLeuLeu.....IleGluCysGlyAlaPro 110
340 GTATGAGGAGATACAAGATCGRAGTCCCTTCGAGTTGGTGTCTTT 389
111 Arg.....SerGlnGlyClnArgSe 117
390 AAATGTCATGTAATGATAATTTATTTGATGTCATACAGGATATGCGTGA 439
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuVal 134
440 AACGCATAACGCCCTCTCGAATATTTATATGATGT..... 476
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
477 .....CCGAATACCTTCGACAGAGTAAGATACATTA 509
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
510 TACAGAGTCAAGCAGAGTTGAATAAAGCAAGGCGATCAATAGCTGGAA 559
159 sMetGlnLysArgAspTyrLysThrAsnValHisPheIleValGlyProp 176
560 GACATCCCTTCAATGATGATGACATCTGAAGTAGAATAATATTATGGCGGAG 609
176 ro.....GlyCysGly 179
610 CATGTTATCGAAGGATTTATTTGGTCTACGGCCCAATGAGGCGGAAGGA 659
180 LysSerLysTrpAlaAlaAsnPheAlaAsnProGluThrThrTyrTrpLy 196
660 AGACAACCTTTGCAAAACATTATTAAGAGACTAAGATGCGTTTATTC 709
196 sPro.ProLysAsnLysTrp 202
```

```
710 GCCAGGAGGAAAAATCATTTGG 729
seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-202-186-17

seq_documentation_block:
; Sequence 17, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/202.186
; APPLICATION NUMBER: US/08/202.186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1104 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-17

alignment_scores:
  Quality: 152.00      Length: 224
  Ratio: 1.267        Gaps: 12
  Percent Similarity: 53.571  Percent Identity: 25.000

alignment_block:
US-09-209-961-3 x US-08-202-186-17 ..
Align seg 1/1 to: US-08-202-186-17 from: 1 to: 1104

17 TtpValPheThrLeuAsnAsnProSerGluAspGluArgLysLysIleAr 33
124 TGGATGTTCCACCATCAACATCCCGCT..... 150
33 gGluLeuProfile.....SerLeuPheAspTyrPheIleValGlyGlu 48
151 .TCGCTACCAGTGATCGGGATGAGTTCAATATATGTTATCATCAAGTGG 199
48 luGlyAsnGluGlyArgThrProHisLeuGlnGlyPheAlaAsnPhe 64
200 AGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGGATACGTCGAGATG 246
65 ValLysLysGlnThrPheAsnLysValLys...TtpTyrLeuGlyAlaAr 80
247 AAAGACGAGAGTCTCTGAAGCAGATGAGAGGCTTCTCCAGGCGCA.. 294
```

```
80 gCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrC 97
   |||:||||| |||:|||||:|||||: |||
295 ....CACCTTGAGAAAGAAAGGGGAGCCAGGAAGACGAGCGGCTTACT 340
   |||:|||||:|||||:|||||:|||||: |||
97 ySerLysGluGlyAsnLeu.....IleGluCysGlyAlaPro 110
   |||:|||||: |||: |||: |||: |||: |||
341 GTATGAAGGAAGATACAAAGTCAAGGTCCTTCGAGTTTGGTCTTT 390
   |||:|||||: |||: |||: |||: |||: |||
111 Arg.....SerGlnGlyGlnArgSe 117
   |||:|||||: |||: |||: |||: |||: |||
391 AATGTCATGTAATGATTAATTTATGTCATACAGGATATCGGTGA 440
   |||:|||||: |||: |||: |||: |||: |||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   |||:|||||: |||: |||: |||: |||: |||
441 AACGCATAAACGGCCTCTGGAATATTTATATGAGTGT..... 477
   |||:|||||: |||: |||: |||: |||: |||
134 hrValAlaGluGlnHisProValThrPheValLysAsn..... 146
   |||:|||||: |||: |||: |||: |||: |||
478 .....CCGAATACCTTCGACAGAAAGTAAGATACATTA 510
   |||:|||||: |||: |||: |||: |||: |||
147 PheArgGlyLeuAlaGluLeuLysVal.....SerGlyLy 159
   |||:|||||: |||: |||: |||: |||: |||
511 TACAGAGTCGACGAGAGTTGAATAAACGAGCGGATGATCTGAA 560
   |||:|||||: |||: |||: |||: |||: |||
159 sMetGlnLysArgAspTrpLysThrAsnValHisPheIleValGlyProp 176
   |||:|||||: |||: |||: |||: |||: |||
561 GACATCCTCAATCATGATGAGCGTCTGAAGTAGAAAATATTATGCGGAGC 610
   |||:|||||: |||: |||: |||: |||: |||
176 ro.....GlyCysGly 179
   |||:|||||: |||: |||: |||: |||: |||
611 CATGTTATCGAAGGATTATTGGTCTTCGCCCAAAATGGAGCGGAAGGA 660
   |||:|||||: |||: |||: |||: |||: |||
180 LysSerLysTrpAlaAlaAsnPheAlaAsnProGluThrTrpTrpLy 196
   |||:|||||: |||: |||: |||: |||: |||
561 AAGCAACGTTTGCAAACATTTAATGAAGACTAAGAAATGCGTTTATTC 710
   |||:|||||: |||: |||: |||: |||: |||
196 sPro.ProLysAsnLysTrp 202
   |||:|||||: |||: |||: |||: |||: |||
711 GCCAGGAGGAAATCAITGG 730
   |||:|||||: |||: |||: |||: |||: |||

seq_name: /cn2_5/ptodata/1/ina/5B_COMB.seq:us-08-202-186-9

seq_documentation_block:
; Sequence 9, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
```

```
TELEFAX: 202 672 5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1111 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: circular
US-08-202-186-9

alignment_scores:
Quality: 147.50 Length: 305
Ratio: 0.952 Gaps: 19
Percent Similarity: 50.820 Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-9
Align seg 1/1 to: US-08-202-186-9 from: 1 to: 1111

17 TrpVal1PheThrLeuAsnProSer.....GluAspGl 28
   |||:|||||: |||: |||: |||: |||: |||
123 TGGATGTTCCACCATCAACATCCCAACACTACCAGTCATGAGGATGA 172
   |||:|||||: |||: |||: |||: |||: |||
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTrpPheIleV 45
   |||:|||||: |||: |||: |||: |||: |||
173 GATATAA.....TATATGGTAT 189
   |||:|||||: |||: |||: |||: |||: |||
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||:|||||: |||: |||: |||: |||: |||
190 ATCAAGTGGAGAGGGGACAGGAGGT...ACTCGTCATGTGCAAGTTAT 236
   |||:|||||: |||: |||: |||: |||: |||
62 AlaasnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
   |||:|||||: |||: |||: |||: |||: |||
237 GTCAGATGAAGAGACGAGCTCTCTGAAGCAGATGAGAGGCTTCTCCC 286
   |||:|||||: |||: |||: |||: |||: |||
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   |||:|||||: |||: |||: |||: |||: |||
287 AGCGCA.....CACCTTGAGAAACGAAAGGAGCCCAAGAAAGAGCGC 330
   |||:|||||: |||: |||: |||: |||: |||
94 ysGluTyrCysSerLysGlu..... 100
   |||:|||||: |||: |||: |||: |||: |||
331 GGTCACTACTGTATGAAGGAAGATACAAAGATCGAGGTCCTTCGAGTTT 380
   |||:|||||: |||: |||: |||: |||: |||
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||:|||||: |||: |||: |||: |||: |||
381 GGTTCATTAAATTTGTCATGT.....AATGA 406
   |||:|||||: |||: |||: |||: |||: |||
117 rAspLeuSerThrAlaValSerThrLeuGluSerGlyIleLeuValT 134
   |||:|||||: |||: |||: |||: |||: |||
407 TAATTTATTTGTCATACAGGATATCGGTGAACG..... 443
   |||:|||||: |||: |||: |||: |||: |||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
   |||:|||||: |||: |||: |||: |||: |||
444 .....CACAAAAGCGCTTTGGAGTATTATATGATTGTCTTAACACC 485
   |||:|||||: |||: |||: |||: |||: |||
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
   |||:|||||: |||: |||: |||: |||: |||
486 TTCGATAGAAGTAAGGATACATTATACAGAGTACAAGCAGAGATGAATA 535
   |||:|||||: |||: |||: |||: |||: |||
162 s.....ArgAspTrpLysThrAsn..... 168
   |||:|||||: |||: |||: |||: |||: |||
536 AACGAAGCGGATGAATAGCTGGAGAACTCTTTCAGTCTGTGGACATCAG 585
   |||:|||||: |||: |||: |||: |||: |||
169 .....ValHisPheIle 172
   |||:|||||: |||: |||: |||: |||: |||
586 AGGTGGAGATATCATGCGCAGCCATGTCATCGGAGATATATTGGGTC 635
   |||:|||||: |||: |||: |||: |||: |||
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
   |||:|||||: |||: |||: |||: |||: |||
636 TATGCCCAAAATGGAGGAGAAAGAACAGCTATGCAAAAACATCTAAT 685
   |||:|||||: |||: |||: |||: |||: |||
```



```

188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTyrTrpAsp. 204
   :::::::::::::::::::: ||| ::::|
686 GAAGACGAGAAATCGTTTATTCT...CCAGGAGGAAATCATTTGGATA 732
205 .....GlyTyrHisGlyGluLysValValValValValValValValVal 218
   ||| ::::| ::::| ::::| ::::| ::::|
733 TATGTAGACTGTATAATACGAGGATATTGTTATTTGAT..... 773
219 GlyTyrLeuPro.....TrpAspAspLeuLeuArgLeuLysCysAspArgTy 233
   ::::| ::::| ::::| ::::| ::::|
774 .....ATTCCAAGATGCAAGAGGATTATTTA.....AATTA 805
233 rProLeuThrValLysGlyGlyThrVal..... 244
   ||| ::::| ::::| ::::| ::::|
806 TGGGTATTAGAGGAATTAGAATGGAATAATTCAAACGGGAATATG 855
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
   ||| ||| ::::| ::::| ::::| ::::|
856 AACCCGTTTGAAGATAGTAGAATATGCGAAGTCATTGTAATGGCTAAC 905
257 GlnThrProLeuGlu 261
   ||| |||
906 TTCCTTCGAGGAA 920

```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-10

```

seq_documentation_block:
; Sequence 10, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-10

```

```

alignment_scores:
  Quality: 147.50      Length: 305
  Ratio: 0.952        Gaps: 19
Percent Similarity: 50.820 Percent Identity: 21.311

```

```

alignment_block:
US-09-209-961-3 x US-08-202-186-10 ..
Align seg 1/1 to: US-08-202-186-10 from: 1 to: 1111
17 TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
   ||| ::::| ::::| ::::| ::::| ::::|
123 TGGATGTTCCACCATCAACAATCCACAACACTACCACTGATGAGGGATGA 172
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   ||| ||| ::::| ::::| ::::|
173 GATCAAA.....TACATGGTAT 189
45 aGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
   ||| ::::| ::::| ::::| ||| ||| ::::| ::::| ::::|
190 ATCAAGTGGAGGGGACACAGGAGGT...ACTCGTCATGTGCAAGGTAT 236
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TriTyrLe 77
   ::::| ::::| ::::| ::::| ::::| ::::|
237 GTCGAGATGAAGAGACGAACTCTCTGAACACAGATGAGAGGCTTCTTCCC 286
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   ||| ||| ::::| ::::| ::::| ::::| ::::|
287 AGCGCA.....CACCTTGAGAAACGAAAGGAGCAAGCAAGAGAGAGCGC 330
94 ySGLuTyrCysSerLysGlu..... 100
   ::::| ::::| ::::| ||| |||
331 GGCATACACTGTATGAAGAGAGATACAAAGATCGAAGTCCCTTCGAGTTT 380
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgse 117
   ||| ::::| ::::| ::::| |||
381 GGTTCAATTAAATGTCATGT.....AATGA 406
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   ::::| ::::| ::::| ::::| ::::|
407 TAATTTATTGATGTCATACAGGATATGCGTGAACG..... 443
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
   ::::| ::::| ::::| ::::| ::::|
444 .....CACAAAGGCCTTTGGAGTATTATATGATTCCTCTTAACACC 485
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
   ||| ::::| ||| ::::| ::::| ::::| ::::|
486 TTCGATAGAAGTAAGATACATTATACAGAGTACACAGCAGACAGATGAATA 535
162 s.....ArgAspTyrLysThrAsn..... 168
   ||| ::::| ::::| ::::| ::::|
536 AACGAAGCGCATGAATAGCTGGAGAACTTCTTTTCAGTGTGGACATCAG 585
169 .....ValHisPheIle 172
   ::::| ::::| ::::| ::::|
586 AAGTGGAGAATATCATGCGCAGCCCATGTCATCGGAGAATAATTTGGGTC 635
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
   ||| ||| ||| ::::| ::::| ::::| ::::|
636 TATGGCCCAATGGAGGAGAAAGAACACACGTATGCAAAACATCTTAAT 685
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTyrTrpAsp. 204
   ::::| ::::| ::::| ||| ::::| |||
686 GAAGACAAGAAATGCGTTTATTCT...CCAGGAGGAAATCATTAGATA 732
205 .....GlyTyrHisGlyGluLysValValValValValValValValVal 218
   ||| ::::| ||| ::::| ::::| ::::|
733 TATGTAGACTGTATAATTACGAAGATATTGTTATTTGAT..... 773
219 GlyTyrLeuPro.....TrpAspAspLeuLeuArgLeuLysCysAspArgTy 233
   ::::| ::::| ::::| |||
774 .....ATTCCAAGATGCAAGAGGATTATTTA.....AATTA 805
233 rProLeuThrValLysLysGlyGlyThrVal..... 244
   ||| ::::| ||| ::::|
806 TGGGTATTAGAGGAATTTAAGAGTGAATAATTCAAACGGGAATATG 855

```



```
;
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-11
```

```
alignment_scores:
  Quality: 144.50      Length: 305
  Ratio: 0.938        Gaps: 19
  Percent Similarity: 50.492  Percent Identity: 21.311

alignment_block:
  US-09-209-961-3 x US-08-202-186-11 ..
  Align seg 1/1 to: US-08-202-186-11 from: 1 to: 1110
  17 TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
  123 TGGATGTTCCACATCAACAAATCCACACACTACGAGTGATGAGGGATGA 172
  28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
  173 GATCAAA.....TATATGGTAT 189
  45 alGlyGluGluGlyAsnGluGlyArgThrProHisLeuGlnGlyPhe 61
  190 ATCAAGTGGAGAGGGACAGGAGGT...ATCGTCATGTCGAAGGTTAT 236
  62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
  237 GTCGAGATGAAGACGAAGCAAGCTCTCTGAAGCAGATGAGAGGCTTCTCC 286
  77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
  287 AGCGGCA.....CACCTTGAGAACCAAGGAAGCAAGCAAGGAAGCGC 330
  94 ysGluTyrCysSerLysGlu..... 100
  331 GGTCACTACTGATGAGGAGATACAGAATCGAAGGTCCTTCGAGTTT 380
```

```
101 GlyAsnLeuLeuIleCysGlyAlaProArgSerGlnGlyGlnArgse 117
  ||| :: :|||
381 GGTGCAATTAAATTGTCATGT.....AATGA 406
  ||| :: :|||
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
  ||| :: :|||
407 TAAATTTATTTGATGTCATACAGGATATGCGTGAACG..... 443
  ||| :: :|||
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
  ||| :: :|||
444 .....CACAAAGGCCTTTGGAGTATTATATATGATTGCTCTTAACACC 485
  ||| :: :|||
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
  ||| :: :|||
486 TTCGATAGAAGTAAGATACATATACAGAGTACAAGCAGACAGATGAATA 535
  ||| :: :|||
162 s.....ArgAspTyrPlyThrAsn..... 168
  ||| :: :|||
536 AACGAAGCGCATGATAGCTGGAGAACTTCTTTCAGTCGATGACATCAG 585
  ||| :: :|||
169 .....ValHisPheIle 172
  ||| :: :|||
586 AGGTGGAATATCATGCGCAGCCATGTCATCGGAGAATAATTGGGTC 635
  ||| :: :|||
173 ValGlyProProGlyCys...GlyLysSerLysTyrAlaAlaAsnPheAl 188
  ||| :: :|||
636 TATGACCAAAATGGAGGAGAAAGACACGATATGCAAAACATCTAAT 685
  ||| :: :|||
188 aAsnProGluThrThrTyrTyrPlyProLysAsnLysTyrTyrPasp. 204
  ||| :: :|||
686 GAAGACGAGAAATGCGTTTATTCT...CCAGGAGAAATCATTTGGATA 732
  ||| :: :|||
205 .....GlyTyrHisGlyGluLysValValIleAspAspPheTyr 218
  ||| :: :|||
733 TATGTAGACTGTATAATTACGAGGATATTGTTATATTGAT..... 773
  ||| :: :|||
219 GlyTyrLeuPro.....TyrAspAspLeuLeuArgLeuCysAspArgTy 233
  ||| :: :|||
774 .....ATCCAGATGCAAGAGGATATTATTA.....AATTA 805
  ||| :: :|||
233 rProLeuThrValLysLysGlyGlyThrVal..... 244
  ||| :: :|||
806 TGGGTATTAGAGGAATTAAAGATGGAATAATTCAAGCGGGAATATG 855
  ||| :: :|||
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
  ||| :: :|||
856 AACCGGTTTGAAGATAGTAGAATAATGTCGAAGTCATTGTAAATGCGTAAC 905
  ||| :: :|||
257 GlnThrProLeuGlu 261
  ||| :: :|||
906 TTTCTTCGGAAGGAA 920
  ||| :: :|||
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-15
seq_documentation_block:
; Sequence 15, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1111 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-15

alignment_scores:
  Quality: 144.50      Length: 305
  Ratio: 0.932        Gaps: 20
  Percent Similarity: 50.820  Percent Identity: 21.967

alignment_block:
US-09-209-961-3 x US-08-202-186-15 ..
Align seq 1/1 to: US-08-202-186-15 from: 1 to: 1111

17  TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
   |||:|||||:|||||:|||||:|||||:|||||:
123 TGGATGTTCCACCACTCAACATCCCAACACTACCAAGTATGAGGGATGA 172
   |||:|||||:|||||:|||||:|||||:|||||:
28  uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
   |||:|||||:|||||:|||||:|||||:|||||:
173 GATCAAA.....TATAGTAT 189

45  alGlyGluGluGlyAsnGluGlyArgThrProHisLeuGlnGlyPhe 61
   |||:|||||:|||||:|||||:|||||:|||||:
190 ATCAAGTGAGAGGGACAGAGGTT...ACTGTCATGTCRAGGTAT 236

62  AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
   |||:|||||:|||||:|||||:|||||:|||||:
237 GTCGAGATGAAGACGCAAGCTCTCTGAAGCAGATGAGAGGCTTCTTCCC 286

77  uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
   |||:|||||:|||||:|||||:|||||:|||||:
287 AGCGCA.....CACCTTGAGAAACGAAAGGGAAGCAAGAGGAGCGC 330

94  ySgLuTyrCysSerLysGlu..... 100
   |||:|||||:|||||:|||||:|||||:|||||:
331 GGTCATCTGTATGAGAGACAGATACAGATCGAAGGTCCTTCGAGTTT 380

101 GlyAsnLeuLeuLeuGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
   |||:|||||:|||||:|||||:|||||:|||||:
381 GGTGCATTAATTTGTCATGT.....AATGA 406

117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
   |||:|||||:|||||:|||||:|||||:|||||:
407 TAATTTATTTGATGTCATACAGGATATGCGTGAACG..... 443

134 hrValAlaGluGlnHisProValThrPheVal.....Lysasn 146
   |||:|||||:|||||:|||||:|||||:|||||:
444 .....CACAAAGGCCCTTGGAGTATTATATGATTGTCCTAAGACC 485

147 Phe...ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
   |||:|||||:|||||:|||||:|||||:|||||:
486 TTCCGATAGAGTAGGATACATATATACAGATACAGCAGAGATGAATAA 535

```

```

162 s.....ArgAspTrpLysThrAsn..... 168
   |||:|||||:|||||:|||||:|||||:|||||:
536 AACGAAGCGGATGAATAGCTGGAGAACTTCTTCAGTCATGGACATCAG 585

169 .....ValHisPheIle 172
   |||:|||||:|||||:|||||:|||||:|||||:
586 AGTGGAGATATATCATGGCGCAGCCATGTCATCGGAGAAATAATTTGGGTC 635

173 ValGlyProGlyCys...GlyLysSerLysTrpAlaIleAsnPheAl 188
   |||:|||||:|||||:|||||:|||||:|||||:
636 TATGCCCAATGGAGGAGAAAGACACGATATGCAAAACATCTAAT 685

188 aAsnProGluThrTrpTyrTrpLysProLysAsnLysTrpTrpAsp. 204
   |||:|||||:|||||:|||||:|||||:|||||:
686 GAAGACGAGAAATCGTTTATCT...CCAGGAGAAATCATTTGGATA 732

205 .....GlyTyrHisGlyLysValValIleAspAspPheTyr 218
   |||:|||||:|||||:|||||:|||||:|||||:
733 TATGTAGACTGTATAATTACGAGGATATTGTATATTGAT..... 773

219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
   |||:|||||:|||||:|||||:|||||:|||||:
774 .....ATTCCAAGATGCAAGAGGATTATTTA.....AATTA 805

233 rProLeuThrValLysThrLysGlyThrVal..... 244
   |||:|||||:|||||:|||||:|||||:|||||:
806 TGGTTATTAGAGGAATTAAGATGGAAATTAATTCAAAGCGGAAATATG 855

245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
   |||:|||||:|||||:|||||:|||||:|||||:
856 AACCGGTTTGAAGATAGTAGAATATGCGAAGTCATTGTAATGCGTAAC 905

257 GlnThrProLeuGlu 261
   |||:|||||:|||||:|||||:|||||:|||||:
906 TTCCTTCGGAAGAA 920

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-13

seq_documentation_block:
; Sequence 13, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:

```

```

; LENGTH: 1109 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
US-08-202-186-13

alignment_scores:
  Quality: 143.50      Length: 368
  Ratio: 0.806        Gaps: 22
  Percent Similarity: 48.370  Percent Identity: 20.652

alignment_block:
US-09-209-961-3 x US-08-202-186-13 ..
Align seg 1/1 to: US-08-202-186-13 from: 1 to: 1109

17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
|||||.....|
123 TGGATGTTTACCATCAACAATCCACACACTACCAGTGATGAGGATGA 172
|||||.....|
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
|||||.....|
173 GATCAAA.....TACATGGTAT 189

45 aGlyGluGluGlyAsnGluGlyArgThrProHisLeuGlnGlyPhe 61
|||||.....|
190 ATCAAGTGGAGGGACAGGAGGT...ACTGTCATGTCAGAGTTAT 236
|||||.....|
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
|||||.....|
237 GTCGAGATGAAGAGCAAGAGCTCTGTAAGCAGATGAGAGGCTTCTTCCC 286
|||||.....|
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
|||||.....|
287 AGGGCA.....CACCTTGAAACGAAGAGGGAGCCCAAGAAGAGCGC 330
|||||.....|
94 yGluTyrCysSerLysGlu..... 100
|||||.....|
331 GGTCACTACTGTGAAGGAAGATCAAGAATCGAAGTCCCTTCGAGTTT 380
|||||.....|
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||||.....|
381 GGTCAATTAAATGTCATG.....AATGA 406
|||||.....|
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
|||||.....|
407 TAATTTATTGATGTCATACAGGATTCGCGTAAACG..... 443
|||||.....|
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
|||||.....|
444 .....CACAAAGGCGCTTGGAGTATTATATGATTGCTTAACACC 485
|||||.....|
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLy 162
|||||.....|
486 TTCGATAGTAAGGATACATATTACAGATACGACGAGATGAATAA 535
|||||.....|
162 s.....ArgAspTrpLysThrAsn..... 168
|||||.....|
536 AACGAAGCGGATGATAGCTGAGAGACTTCTTTTCAGTCTTGGACATCAG 585
|||||.....|
169 .....ValHisPheIle 172
|||||.....|
586 AGGTGGAGATATCATGCGCGCAGCATGTCATCGGAGAATAATTGGTGC 635
|||||.....|
173 ValGlyProProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
|||||.....|
636 TATGCCCAATAGGAGGAAGAAAGAACACAGCTATGCAAAACGCTTAAT 685
|||||.....|
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
|||||.....|
686 GAAGACGAGAATCGGTTTATCT...CCAGGAGAAATCATTTGGATA 732
|||||.....|

```

```

205 .....GlyTyrHisGlyGluLysValValValIleAspPheTyr 218
|||||.....|
733 TATGTAGACTGTATAATTACGAGGATATTGTTATATTGAT..... 773
|||||.....|
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||||.....|
774 .....ATTCCAAGATGCAAGAGGATATTATTA.....AATTA 805
|||||.....|
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|||||.....|
806 TGGGTATTAGAGGAATTGAAGATGGAATAATTCAAAGCGGGAATATG 855
|||||.....|
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
|||||.....|
856 AACCCGTTTGAAGATAGTAGATATATGTCGAAGTCATTGTTAATGGCTAAC 905
|||||.....|
257 GlnThrProLeuGluTrpTyrSerSer..... 265
|||||.....|
906 TTCCTTCGAGGAAGGAATCTTTTCTGAAGATCGAATAAAGTTGGTTTC 955
|||||.....|
266 .....ThrAlaValProAlaValGluAlaLeuT 275
|||||.....|
956 TTGCTGAACAAGTAATGACTTTTACAGCGCAGCTCCGACAAAAGCGCACT 1005
|||||.....|
275 Yr.....ArgArgIleThrSerLeuValPheTrpLysAsnAlaThrLys 289
|||||.....|
1006 ATGACAAAAGACAGCTGCTGATTGACATCTGACAGCATCTAGGCGCGTA 1055
|||||.....|
290 GlnSerThrGluGluGlyGlnPheValThrLeuSerProCysPr 306
|||||.....|
1056 GGCGGTGAGCAATGAACGCGGAGATCATA.....TGTC 1090
|||||.....|
306 oGlu 307
|||||
1091 CGAG 1094

seq_name: /cgn2_5/ptodata/1/ins/5B_COMB.seq:US-08-202-186-12

seq_documentation_block:
; Sequence 12, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:

```

```
: LENGTH: 1111 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: circular
US-08-202-186-12

alignment_scores:
  Quality: 143.50      Length: 305
  Ratio: 0.932        Gaps: 19
  Percent Similarity: 50.492  Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-12
..
Align seg 1/1 to: US-08-202-186-12 from: 1 to: 1111

17 TrpValPheThrLeuAsnProSer.....GluAspG1 28
123 TGGATGTTCCATCAACAATCCCAACAACACTACCAGTGATGAGGATGA 172
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
173 GATCAAA.....TATATGGTAT 189
45 alGlyGluGluGlyAsnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
190 ATCAAGTGGAGAGGGACAGAGGGT...ACTGCTCATGTGCAAGGATAT 236
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TrpTyrLe 77
237 GTCGAGATGAAGACGAGAGCTCTCTGAAGCAGATGAGAGCTTCTTCCC 286
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGluAsnL 94
287 AGCGCA.....CACCTTGAGAAACGAAGGGAAGCAAGCAAGAGCGC 330
94 yGluTyrCysSerLysGlu..... 100
331 GGTCACTACTGTATGAGGAGATACAGAAATCGAAGTCCTTCGAGTTT 380
101 GlyAsnLeuLeuLeuGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
381 GGTGCTTTAAATGTCTATGT.....AATGA 406
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
407 TAATTTATTGATGTCATACAGATATCGGTGAACG..... 443
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
444 .....CACAAAGCCCTCTCGAGTATTATGATTCCTCTAACACC 485
148 .....ArgGlyLeuAlaGluLeuLeuLysValSerGlyLysMetGlnLy 162
486 TTCGATAGAAGTAAAGCATACATTATACAGAGTACAGCAGACATGAATA 535
162 s.....ArgAspTrpLysThrAsn..... 168
536 AACGAAGCGCATGAATAGTGGAGAACGCTTTCAGTGCCTTGGACATCAG 585
169 .....ValHisPheIle 172
586 AAGTGGAGAATATCATGGCGCAGCCATGTCATCGGAGAATAATTGGGTC 635
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
636 TATGGCCCAATCGAGGAGNAGAGACACACGATGTCACAAACAATAAT 685
188 aAsnProGluThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
686 GAAGACGAGGAATGCGTTTATTCT...CCAGGGGGAATAATCATTCGATA 732

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-202-186-14
seq_documentation_block:
; Sequence 14, Application US/08202186
; Patent No. 5756708
; GENERAL INFORMATION:
; APPLICANT: KARAN, Mirko
; APPLICANT: BURNS, Thomas M.
; APPLICANT: DALE, James L.
; APPLICANT: HARDING, Robert M.
; TITLE OF INVENTION: DNA Sequences of Banana Bunchy Top Virus
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/202,186
; FILING DATE: 24-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: JEFFERY, Donald D.
; REGISTRATION NUMBER: 19,980
; REFERENCE/DOCKET NUMBER: 71611/102 FIKE
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 672 5300
; TELEFAX: 202 672 5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1110 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: circular
; US-08-202-186-14

alignment_scores:
  Quality: 140.50      Length: 305
  Ratio: 0.918        Gaps: 19
  Percent Similarity: 50.164  Percent Identity: 21.311

alignment_block:
US-09-209-961-3 x US-08-202-186-14
..
Align seg 1/1 to: US-08-202-186-14 from: 1 to: 1110
```

```
17 TrpValPheThrLeuAsnAsnProSer.....GluAspG1 28
|||||.....|
122 TGGAGTTTACCATCAACATCCACACTACCAGTCATGAGGACGA 171
|||||.....|
28 uArgLysLysIleArgGluLeuProIleSerLeuPheAspTyrPheIleV 45
|||.....|
172 GATCAA.....TACATGGTAT 188
|||.....|
45 alGlyGluGluCysGlnGluGluGlyArgThrProHisLeuGlnGlyPhe 61
|||.....|
189 ATCAAGTCGAGGGGACAGGAGGCT...ACTCGTCATGTGCAAGGATAC 235
|||.....|
62 AlaAsnPheValLysLysGlnThrPheAsnLysValLys...TyrTyrIle 77
|||.....|
236 GTGGAGATGAGAGGACGAGGAGCTCTGAGACAGATGAGAGGCTTCTCC 285
|||.....|
77 uGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGlnAsnL 94
|||.....|
286 AGGCCA.....CACCTTGAGAACGAAAGGGGGCCCAAGATGAGCGC 329
|||.....|
94 ysGluTyrCysSerLysGlu..... 100
|||.....|
330 GGTCACTACTGATGAGGAGGATCAAGAAATCGAAGGTCCTTCGAGTTT 379
|||.....|
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
|||.....|
380 GGTGCATTAAATGTCATGT.....AATGA 405
|||.....|
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValT 134
|||.....|
406 TAATTATTGTGTCATACAGGATGCGTGCAAGC..... 442
|||.....|
134 hrValAlaGluGlnHisProValThrPheValLysAsnPhe..... 147
|||.....|
443 .....CACAAAGACCTTTGGAGTATTTATGATGTCCTTAATACC 484
|||.....|
148 .....ArgGlyLeuAlaGluLeuLysValSerGlyLysMetGlnLys 162
|||.....|
485 TTCGATGAAGTAAGGATACATTATACAGATGACAAAGCAGAAATGAATA 534
|||.....|
162 s.....ArgAspTrpLysThrAsn..... 168
|||.....|
535 AACGAGGCGGATGATAGCTGGAGAACGCTTTCAGTCTCGACATCAG 584
|||.....|
169 .....ValHisPheIle 172
|||.....|
585 AGGTTGAGATATCATGCGCGACCATGTCATCGAAGATATTATTGGGT 634
|||.....|
173 ValGlyProGlyCys...GlyLysSerLysTrpAlaAlaAsnPheAl 188
|||.....|
635 TACGCCCAATGGAGGAGGAGGAAAGACACGATGCAAAACATCTAAT 684
|||.....|
188 asnProGlnThrThrTyrTrpLysProProLysAsnLysTrpTrpAsp. 204
|||.....|
685 GAAGACGAGAAATGCGTTTATCT...CCAGGAGGAAATCATTGGATA 731
|||.....|
205 .....GlyTyrHisGlyGluLysValValValIleAspAspPheTyr 218
|||.....|
732 TATGTAGACTGTATTAATATGAGGATATTGTTATATTGAT..... 772
|||.....|
219 GlyTrpLeuPro.....TrpAspAspLeuLeuArgLeuCysAspArgTy 233
|||.....|
773 .....ATCCCTAGATGCAAGAGGATATTATTA.....AATTA 804
|||.....|
233 rProLeuThrValLysThrLysGlyGlyThrVal..... 244
|||.....|
805 TGGTTTATTAGAGGAATTAAAGATGGAATAATTCAAAAGCGGAAATATG 854
|||.....|
245 ..ProPheLeuAla.....ArgSerIleLeuIleThrSerAsn 256
|||.....|
855 AACCCGTTTGAAGATGTAGAAATATGTCGAAGTCAATTGTAATGCGTAA 904
|||.....|
```

```
257 GlnThrProLeuGlu 261
||| |||
905 TTCCCTCCGAGGAA 919

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-07-674-852-1

seq_documentation_block:
: Sequence 1, Application US/07674852
: Patent No. 5476658
: GENERAL INFORMATION:
: APPLICANT: TSAREV, SERGEI A
: APPLICANT: EMERSON, SUZANNE U
: APPLICANT: BALAYAN, MICHAEL S
: APPLICANT: PURCELL, ROBERT H
: TITLE OF INVENTION: A VACCINE AGAINST HEPATITIS A
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: CUSHMAN, DARBY & CUSHMAN
: STREET: Eleventh Floor, 1615 L. Street, N.W.
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20036-5601
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25;
: SOFTWARE: Word Perfect 5.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/674,852
: FILING DATE: 19910326
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: SCOTT, WATSON T
: REGISTRATION NUMBER: 26,581
: REFERENCE/DOCKET NUMBER: WTS/5683/83919/SRL
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202) 861-3000
: TELEFAX: (202) 822-0944
: TELEX: 6714627CUSH
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7400 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: US-07-674-852-1

alignment_scores:
Quality: 111.00 Length: 206
Ratio: 1.156 Gaps: 12
Percent Similarity: 46.602 Percent Identity: 25.728

alignment_block:
US-09-209-961-3 x US-07-674-852-1 ..

Align seg 1/1 to: US-07-674-852-1 from: 1 to: 7400

138 GlnHisProValThrPheValLysAsnPheArgGlyLeuAlaGluLeuLe 154
||| ||| .....|
4180 CAGAAGGAGGTGAACTCATTCAGAAATTCAGAACACAGTTCATCCCTGC 4229
||| .....|
154 uLysValSerGlyLysMetGlnLys.....ArgAspTrpL 166
||| .....|
4230 CCAGGTCGACTCTAGTTTGTCTCATTGTCCACCACCTGAGAGATTGTA 4279
||| .....|
166 ysThrAsnValHis..... 170
||| .....|
4280 TTGCTAGATCCATCAAAACATTAAGAAATTTAGGTCATTAATCAGGCT 4329
||| .....|
171 .....PheIleValGlyProProG1 177
```

```
4330 ATGCTGACTAGGTGTAACCTGGTCTGTATTATATATGTAAGAGAGG 4379
177 yCysGlySerLysTrpAlaAlaAsnPheAla..... 188
4380 TGGAGGAAGAGTTTAACCTCTATTGCTTGGCAACAAAAATTTGCCAAC 4429
189 .....AsnProGlu...ThrThrTyrTrpLysProProLysAsn 200
4430 ATATGCTGTGTGAACCAAGAAAGATATATATACAAACCTGTGCTTCA 4479
201 LysTrpTrpaspGlyTyrHisGlyLysValValValIleAspaspPh 217
4480 GACTACTGGGATGATAGTAGTGGTCAATTGGTTGTATCATTTGATGACAT 4529
217 eTyrGlyTrp.....LeuProTyrAspAspLeuLeuArgLeuC 230
4530 T...GGTCAAAATACTACAGATGAAGATTGTCAGATTTTGTCAATTGG 4576
230 ysAspArgTyrProLeu.....ThrValLysThrLysGly 241
4577 TGCTGCTTGTCTATGAGGTAAATATGCTTCTTTGGAGAGAAAGGG 4626
242 Gly.....ThrValProPheLeuAlaArgSerIleLeuIleThrSerAs 256
4627 AGACACTTTTCTCCCGGTTT.....ATAATTGCCACATCAAA 4664
256 nGlnThrProLeuGluTrpTyr.....SerSerThrAlaValProAlav 271
4665 T.....TGGTCAATCCCAAGTCTCTTAAGACTGTTTATGTGA 4699
271 alGluAlaLeuTyrArgArgIleThrSerLeuVal..... 282
4700 AGGAGCTATAGATCGCGCTTCATTATAGATTGAAGTCAAAACCAGCA 4749
283 ...PheTrpLysAsnAla 287
4750 TCCTTTTACAAAAATGCA 4767
```


OM of: US-09-209-961-3 to: EST:* out_format : pfs

Date: Dec 23, 1999 5:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame p2n.model -DEV=xlp
-O=/cgn2.1/USPTO.spool/US09209961/runat_22121999_101043_17901/app_query.fasta.1
-DB=EST -OFT=fastap -SUFFIX=1st -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -GAPOP=6.000
-GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -LOCALIGN=200 -THR_SCORE=escore -ALIGN=3 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-3

Query length: 314

Database: EST:*

Database sequences: 4089388

Database length: 1713965092

Search time (sec): 1572.090000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est21:AA906356	- 92.50	175.96	0.4420	716	AA906356 0194c11.s1 Soares_NFL
gb_est24:AA198072	+ 86.50	171.59	0.7750	346	AA198072 q154a05.x1 NCI_CGAP_B
gb_est10:AA181774	+ 87.50	171.38	0.7959	422	AA181774 zp54d01.s1 Stratagene
gb_est1:T55734	+ 87.50	170.98	0.8374	436	T55734 yb40c11.s1 Stratagene fe
gb_est8:AA022948	+ 87.50	168.94	1.09	516	AA022948 ze72h01.s1 Soares.feta
gb_est21:AA936814	+ 85.50	167.48	1.31	405	AA936814 cb89c05.s1 NCI_CGAP_C
gb_est29:AA159903	+ 83.50	166.92	1.41	295	AA159903 tmlg02.x1 NCI_CGAP_B
gb_est2:T74565	+ 85.00	166.29	1.53	408	T74565 yc36d10.s1 Stratagene 11
gb_est5:AA523218	+ 86.00	165.50	1.69	522	AA523218 ni44h07.s1 NCI_CGAP_Lu
gb_gsl1:CMS00127	- 89.50	164.15	2.01	1101	AL078747 Drosophila melanogast
gb_est1:AA226908	+ 83.50	164.11	2.02	372	AA226908 zr19e04.s1 Stratagene
gb_est7:W20432	+ 83.50	163.89	2.08	379	W20432 zc58c05.s1 Soares.fetal
gb_gss7:AA0469737	- 85.00	163.71	2.13	505	AA0469737 CITB1-EI-2584G10.RR CI
gb_est11:AA218011	+ 86.50	163.63	2.15	367	AA218011 mv59e06.r1 Soares.mous
gb_est5:H95640	+ 83.50	163.57	2.17	389	H95640 yv17a07.s1 Soares.fetal
gb_est24:AA2121630	+ 83.50	162.96	2.34	409	AA2121630 qg20f10.x1 Soares.plac
gb_est11:AA243218	+ 83.50	162.85	2.38	413	AA243218 zt26g03.s1 Stratagene
gb_est15:AA356673	+ 83.00	162.44	2.50	390	AA356673 EST177 Sugarcane leaf
gb_est17:AA653206	+ 83.50	162.24	2.57	434	AA653206 ag64e11.s1 Gessler.Wil
gb_est9:AA070716	+ 83.50	162.19	2.59	436	AA070716 zm66f02.s1 Stratagene
gb_est10:AA179708	+ 83.50	162.16	2.60	437	AA179708 zp53b02.s1 Stratagene
gb_est10:AA179708	+ 83.50	162.08	2.62	440	AA179708 zp53b07.s1 Stratagene
gb_est11:AA205826	+ 83.50	162.08	2.62	440	AA205826 zq53h03.s1 Stratagene
gb_est20:C85851	- 86.00	161.90	2.68	586	C85851 C85851 Mouse fertilized
gb_est38:AA057497	- 86.00	161.67	2.76	716	AA057497 ca04g01.x1 C.elegans f
gb_est9:AA070466	+ 83.50	161.41	2.86	465	AA070466 zm70b08.s1 Stratagene
gb_est5:H94863	+ 84.50	161.35	2.88	560	H94863 yu57d04.s1 Soares.fetal
gb_est11:AA227144	+ 83.50	161.33	2.89	468	AA227144 zr22c01.s1 Stratagene
gb_est29:AA168753	+ 83.50	161.20	2.94	473	AA168753 zt90b09.x1 NCI_CGAP_HN
gb_est22:AA1006723	- 84.50	161.05	2.99	574	AA1006723 ua67g09.r1 Soares.2NDM
gb_est31:AA1688536	+ 83.50	161.00	3.01	481	AA1688536 wc90d06.x1 NCI_CGAP_C
gb_est9:AA070410	+ 83.50	160.82	3.08	488	AA070410 zm68b12.s1 Stratagene
gb_est26:AA1371057	+ 83.50	160.72	3.12	492	AA1371057 ta29g10.x1 Soares.feta
gb_est1:T56604	+ 84.50	160.52	3.21	600	T56604 yb35b03.s1 Stratagene fe
gb_est23:AA135508	+ 84.00	160.28	3.31	559	AA135508 GH13304.5prime GH Dros
gb_est11:AA216784	+ 83.50	160.26	3.31	511	AA216784 zq46f10.s1 Stratagene
gb_est2:AA858195	+ 83.50	160.12	3.37	517	AA858195 oe43b05.s1 NCI_CGAP_P
gb_est5:H79483	+ 82.50	159.99	3.43	436	H79483 yu49b04.s1 Soares.fetal
gb_est1:T15728	+ 83.50	159.35	3.72	551	T15728 IB1798 infant brain, Bed
gb_est31:F30063	- 80.50	159.34	3.73	320	F30063 HSPD20323 HM3 Homo sapie
gb_est18:AA706656	+ 83.50	159.31	3.74	553	AA706656 ag90h01.r1 Stratagene
gb_est11:AA206124	+ 83.50	159.28	3.75	554	AA206124 zq52g12.s1 Stratagene

gb_est21:AA906004 + 82.50 159.16 3.82 467 ! AA906004 0j84b06.s1 Soares_N
gb_est4:D51888 - 80.50 159.15 3.82 325 ! D51888 HM056A12B Clontech h
gb_est19:AA778101 + 83.50 159.13 3.83 561 ! AA778101 zf45g04.s1 Soares_f

seq_name: gb_est21:AA906356

seq_documentation_block:
LOCUS AA906356 716 bp mRNA EST 19-MAY-1998
DEFINITION OJ94c11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1505972 3' similar to gb:X05360 CELL DIVISION CONTROL PROTEIN
2 HOMOLOG (HUMAN);, mRNA sequence.

ACCESSION AA906356

NID 93041479

VERSION AA906356.1 GI:3041479

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 716)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished. (1997)

COMMENT On Jan 14, 1998 this sequence version replaced gi:1797807.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Trace considered overall poor quality

Insert Length: 1211 Std Error: 0.00

Seq primer: -40m13 fwd. Et from Amersham

High quality sequence stop: 1.

FEATURES

Source

Location/Qualifiers

1..716

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1505972"

/lab_host="DH10B"

/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site: 1: Not 1; Site: 2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung NBHL19W, testis NHT, and B-cell

NCI_CGAP_GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297480-302087, 682632-687239,

726408-728711, and 729096-731399. Subtraction by Bento

Soares and M. Fatima Bonaldo.

BASE COUNT 207 a 156 c 162 g 191 t

ORIGIN

alignment_scores:

Quality: 92.50 Length: 193

Ratio: 1.051 Gaps: 9

Percent Similarity: 45.596 Percent Identity: 20.725

alignment_block:

US-09-209-961-3 x AA906356/rev ..

Align seg 1/1 to reverse of: AA906356 from: 1 to: 716

92 GlnAsnLysGluTyrCysSer.....LysGluGlyAs 102

688 GAATATCCGAGAGTGCACAAATTCGATCCCGAGTGAAGTGCATC 639

102 nLeuLeulle.....GluCysGlyAlaProArgSerGlnGlyGln.... 115

```

|||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : : |||||: : : : :
638 TCAGTCTAGTACGCAATGTGGTGCCTATACCTCCAGTGTCCAGTGT 589
116 ..... ArgSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGly 130
588 AGTTGAGTACCGAGTACCCAGCAATCTCTCTGAAGATCCGTACCGGA 539
131 IleLeuValThrValAlaGluGlnHisProValThrPheValLysAsnPh 147
538 GTGTACTACTCATGTGCATACACACGAGTGGTATCCAAAAGCTCCT 489
147 eArgGlyLeuAlaGluLeu..... 153
488 GCAAGGCCAATGACGCACTTTAATTTCTCTGAGTCATCAATCAGGAG 439
154 ..LeuLysValSerGlyLysMetGlnLysArgAspTrpLysThrAsnVal 169
438 ATTTGAAGGTACAAAGTCTCTGTGACGAACCTTCTGGAGGAACAACA 389
170 HisPheIleValGly.....ProPr 176
388 CATTCCTCCTAGGATTGGTATAGATACTCTTAACAAGTGAGGAATCC 339
176 oGlyCysGlySerLysTrpAlaLysPheAlaAsnProGluThrT 193
338 ATGTACTGCCAGGAGGATGATCCAAAGTATTTCTCGAGATCCA... 293
193 hrTyrTrpLysProProLysAsn..... 200
292 .....TGAAGGAACCTCAAGATGAGATATACCTCGAATCCTGTATAG 248
201 .....LysTrpTrpAsp.....G1 205
247 CAACACCTTGAGGACTGACTATATGTGTATGACGAGTTCCTTTAATAG 198
205 yTyrHisGlyGlu.....LysValValLysIleA 215
197 GAAATTTCCGGAATTCAGTATTAGGACCCCTTTCGTCTACACTGTCTAG 148
215 spAspPheTyrGlyTrpLeuProTrpAsp 224
147 ACTGATTTTGTTCAGGCTACCACTCGAC 119

```

seq_name: gb_est24: A1198072

```

seq_documentation_block: 346 bp mRNA EST 02-DEC-1998
LOCUS A1198072
DEFINITION q154a05.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:1860272 3'
similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN (HUMAN);, mRNA
sequence.
ACCESSION A1198072
NID 93750678
VERSION A1198072.1 GI:3750678
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 346)
AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute / National Institute of Neurological
Disorders and Stroke, Brain Tumor Genome Anatomy Project
(CGAP/BTGP), Tumor Gene Index
JOURNAL Unpublished (1998)
COMMENT On Aug 21, 1998 this sequence version replaced.

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 771 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 217.

FEATURES

source
1..346
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="699EL2: 8"
/clone_lib="NCI_CGAP_Brn25"
/tissue_type="anaplastic oligodendroglioma"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGGAGCGCGCATAGTGTGTGTGTGTGTGTGTGT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library is normalized, and was constructed by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT 107 a 67 c 82 g 89 t 1 others
ORIGIN

alignment_scores:

Quality: 86.50 Length: 127
Ratio: 1.602 Gaps: 6
Percent Similarity: 42.520 Percent Identity: 25.984

alignment_block:

US-09-209-961-3 x A1198072

Align seg 1/1 to: A1198072 from: 1 to: 346

```

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
24 TGGCTCTTGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAG 70
87 ..... 87
71 GTTCAACACAAAATAGAAATCTCAATGTAGTAGATAGACAAAACCAAGTG 120
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
121 TGTGAGGGGGAACACAGCAAGAAAGGAAGATGATGTTCAGAAAAA 170
100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGln 116
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
171 GATGAGGA.....GGTTCCCTCTCTCTGGGACTGA 205
116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeu 133
|||::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
206 CTCAAACTGATGTGCGAGTATACACCATCCAGAGTCAGGGGTGTTC 255
133 alThrValAlaGluGlnHisProValThrPheValLysAsnPheArgGly 149
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
256 TTCTT.....TTTGTGGGA 269
150 LeuAlaGluLeuLysValSerGlyLysMetGlnLysArgAspTrpLys 166
::: ||||| ||||| ||||| ||||| ||||| ||||| |||||
270 .....GTAAGAAAGGTGGGATTAATAAAACCGTTTCTCGAG 307
166 sThrAsnValHisPheIleValGlyProPro 176
: : : : : |||||
308 GCTTAGGACCAAGCTTGTCTCTTCCCTCCCT 338

```

seq_name: gb_est10:AAL81774

```

seq_documentation_block: 422 bp mRNA EST 06-JAN-1997
LOCUS AAL81774
DEFINITION zp54001.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
cDNA clone IMAGE:613249 3' similar to gb:J00314_rna2 TUBULIN BETA-1
CHAIN (HUMAN); mRNA sequence.
ACCESSION AAL81774
NID g1765241
VERSION AAL81774.1 GI:1765241
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 422)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.
TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On May 9, 1995 this sequence version replaced gi:804184.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 312.

```

```

FEATURES
source
1..422
/organism="Homo sapiens"
/db_xref="GDB:4645088"
/db_xref="taxon:9606"
/clone="IMAGE:613249"
/tissue_type="neuroepithelial cells"
/lab_host="Stratagene NT2 neuronal precursor 937230"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT 126 a 77 c 109 g 109 t 1 others
ORIGIN

```

```

alignment_scores:
Quality: 87.50 Length: 100
Ratio: 1.902 Caps: 4
Percent Similarity: 46.000 Percent Identity: 30.000

```

```

alignment_block:
US-09-209-961-3 x AAL81774 ..
Align seg 1/1 to: AAL81774 from: 1 to: 422

```

```

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
|||||
34 TGGCTCTTGGG...CGATGTCATCTTTTCATATGAAAGAGCAAA 80
87 ..... 87

```

```

81 GTTCAACACAAAATAGAAATCTCAAAATGTAGGATAGACAAACCAAGTG 130
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
||| |||||:||||| |||||:|||||
131 TGTGAGGGGGGAGCAACAGCAAGAAAGAAATGAGATGTTTCCAAAAA 180
100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
:||||: ||| ||| |||
181 GATGGAGGA.....GGTTCCCTCTCCTCTGGGACTGA 215
116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
|||||: |||||: |||||: |||||: |||||: |||||: |||||:
216 CTCAAACACTGTTGGCAGTATACACCATTCAGAGTCAGGGGTGTCTCA 265
133 alThrValalagluGlnHisProValThrPheValLysAsnPheArgGly 149
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
266 TTCTTTTGGGAGTAAGAAAGGTGGGATTAAAGAGCGTTTCTGGGA 315
seq_name: gb_est1:T55734

```

```

seq_documentation_block:
LOCUS T55734 436 bp mRNA EST 06-FEB-1995
DEFINITION yb40c11.s1 Stratagene fetal spleen (#937205) Homo sapiens cDNA
clone IMAGE:73652 3' similar to gb:J00314_rna2 TUBULIN
BETA-1 CHAIN (HUMAN), mRNA sequence.
ACCESSION T55734
NID 9657595
VERSION T55734.1 GI:657595
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 436)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.
and Marra,M.

```

```

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 3177
High quality sequence stops: 275 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the IMAGE
Consortium (info@image.llnl.gov) for further information.
Insert Length: 3177 Std Error: 0.00
Seq primer: -2lm13
High quality sequence stop: 275.

```

```

FEATURES
source
1..436
/organism="Homo sapiens"
/db_xref="GDB:495317"
/db_xref="taxon:9606"
/clone="IMAGE:73652"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal"
/dev_stage="fetal"
/note="Organ: spleen; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG

```

```

Location/Qualifiers
1..436
/organism="Homo sapiens"
/db_xref="GDB:495317"
/db_xref="taxon:9606"
/clone="IMAGE:73652"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal"
/dev_stage="fetal"

```

```

FEATURES
source
1..436
/organism="Homo sapiens"
/db_xref="GDB:495317"
/db_xref="taxon:9606"
/clone="IMAGE:73652"
/clone_lib="Stratagene fetal spleen (#937205)"
/tissue_type="fetal"
/dev_stage="fetal"
/note="Organ: spleen; Vector: pBluescript SK-; Site_1:
EcORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Pooled spleens. Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGG

```

```

BASE COUNT      122 a      83 c     110 g     112 t       9 others
ORIGIN

alignment_scores:
    Quality:   87.50          Length:   117
    Ratio:     1.750         Gaps:      6
    Percent Similarity: 42.735    Percent Identity: 26.496

alignment_block:
US-09-209-961-3 x T55734 ..

Align seg 1/1 to: T55734 from: 1 to: 436

75  TrpTyrLeuGlyAlaAArgCysHisIleuLysAlaLys..... 87
||||| ||||| ||||| ||||| : :: |||
30  TGGCTCTTTGGG...CGATGTCACTCTTTTCATATGAAAAAACACCAA 76
87 ..... 87

77  GTTCAACACAAAATAGAAATCTCAATGTAGGATAGAACAACCAAGTN 126

88  .... GlyThrAspGlnAsnLysGlu..... TyrCysSerLys 99
||||| ||||| ||||| ||||| | |||||
127 TNTGAGGGGGGAAGCACACAGCAAAAGGAAGAAATGCAGATN TTGCCAAAA 176

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
:::|||::: ||| ||| |||
177 GATGGAGGA.....GGGTCCCTCTCTCTCGGGGACTGA 211

116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
||||: ::::: ||||| |||::: ||||| ||||| :::::
212 CTCAAACACTGATGTGGCAGTATACACCATTCCAGAGTCAGGGGNTTCA 261

133 alThrValAlaGluGlnHisProValThrPheValLysAsnPheArgGly 149
:: ||| ||| |||
262 TTCTT.....TTTTTGGA 275

150 LeuAlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLy 166
::::: ::::: ||| ||| ::::: ||| ||||
276 .....GTAAGAAAGGTGGGGGNTTAAGAAGACGTTTCTCGAA 313

166 s 166
|
314 G 314

seq_name: gb_est8:AA022948

seq_documentation_block:
LOCUS AA022948 516 bp mRNA EST 30-JAN-1997
DEFINITION ze72h01.s1 Soares fetal heart.NBH19W Homo sapiens CDNA clone
IMAGE:364561.3. similar to gb:300314_rna2 TOBULIN BETA-1 CHAIN
(HUMAN);, mRNA sequence.

ACCESSION AA022948
NID gl487038
VERSION AA022948.1 GI:1487038
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
1 (bases 1 to 516)
AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le.M., Le.N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Rohlfing,T., Scheilenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Trevasakis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R.,
and Marra,M.

Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
'EDLINE' 97044478
```

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 798 Std Error: 0.00
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 212

FEATURES

source

```

1. 516
/organism="Homo sapiens"
/db_xref="GDB:1281265"
/db_xref="taxon:9606"
/clone="IMAGE:364561"
/clone_lib="Soares_fetal_heart_NBHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; vector: pT73D (Phar
modified polylinker; Site_1: Not I; Site_
strand cDNA was primed with a Not I - oli
TGTACCAATCTGAAGTGGAGCGGCATCTTTTTTTT
double-stranded cDNA was size selected, l
adapters (Pharmacia), digested with Not I
the Not I and Eco RI sites of a modified
(Pharmacia). Library went through one rou
normalization to a Cot = 5. Library const
M.Fatima Bonaldo. This library was constr
same fetus as the fetal lung library, Soa
NBHL19W."
160 a 102 c 134 g 116 t 4 others
BASE COUNT
ORIGIN

```

```
alignment_scores:
  Quality: 87.50      Length: 187
  Ratio: 1.389       Gaps: 9
  Percent Similarity: 33.690  Percent Identity: 22.460

alignment_block:
  US-09-209-961-3 x AA022948      .
  Align seq 1/1 to: AA022948      from: 1 to: 516
```

```

75 TrpTrpLeuGlyAlaArgCysHisIleGluLysAlaLys.....87
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
18 TGCGCTCTGGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAA 64
   ..
87 .....87
65 GTTCAACACAAAAATAGAAATCTCAAATGTAGGATAGAAACAAACCAAGTG 114
   ..
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
   ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 TGTAGCGGGGGAACCAACGAAAAAGGAAGAAATGAGATGTTGCAAAAAA 164
   ..
100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
   :|||: ||| ||| |||
165 GATGGAGGA.....GGGTCCCTCTCCCTCGGGGACTGA 199
   ..
116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
   ||||: :|||: ||||| ||||| ||||| ||||| ||||| :|||:
200 CTCAAACACTGATGTGGCAGTATACACCATCCAGAGTCAGGGGTGTTCA 249
   ..
133 alThrValAlaGluGlnHisProValThrPheValLysAsnPheArgGly 149
   : : |||
250 TCTCTTTTTCGGAA.....263
   ..
150 LeuAlaGluLeuLysValSerGlyLysMetGlnLysArgAspIleV 164

```

RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo.

BASE COUNT 126 a 75 c 107 g 97 t
ORIGIN

alignment_scores:
Quality: 85.50 Length: 100
Ratio: 1.819 Gaps: 4
Percent similarity: 47.000 Percent Identity: 29.000

alignment_block:

US-09-209-961-3 x AA936814

Align seg 1/1 to: AA936814 from: 1 to: 405

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87

18 TGGCTCTTGGGG...CGATGTCATCTTTCAATATGAAAAAAGCAGCAA 64

87 87

65 GTTCAACACAAATAGAAATCTCAATGTAGGATAGACAAACCAAGTG 114

88GlyThrAspGlnGlnAsnLysGluTyr.....CysSerLys 99

115 TGTGACGGGGGAAGCAACAGCAAAAGGAAGAAATGAGATGTTGCAAAAA 164

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116

165 GATGGAGGA.....GGTTCCCTCTCTCTGGGGACTGA 199

116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133

200 CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTCA 249

133 alThrValAlaGluGlnHisProValThrPheValLysAsnPheArgGly 149

250 TTCTTTTGGGAGTAAGAAAGGTGGGATTAAATAAAACGTTCTTAGGA 299

seq_name: gb_est29:AI589903

seq_documentation_block:

LOCUS AI589903 295 bp mRNA EST 21-APR-1999
DEFINITION tm81g02.x1 NCI_CGAP_Brn25 Homo sapiens cDNA clone IMAGE:2164562 3', similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN (HUMAN);, mRNA

SEQUENCE

ACCESSION AI589903

NID 94598951

VERSION AI589903.1 GI:4598951

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 295)

AUTHORS NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute / National Institute of Neurological

Disorders and Stroke, Brain Tumor Genome Anatomy Project

(CGAP/BTGP), Tumor Gene Index

JOURNAL Unpublished (1998)

COMMENT On Mar 10, 1998 this sequence version replaced gi:2948781.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: David N. Louis, M.D., Myrna R. Rosenfeld M.D.,

Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

264 |||

166 sThrAsnValHisPheIleValGlyProGlyCysGlyLysSerLysT 183

281 ATTAAAGAACCGTTT.....CTGGAGGCTTAGGGACCAAGGCCT 321

183 rPalalaAAsnPheAlaAsnPro..... 190

322 GGTCTCTT...TTCCCTCTCCCAACCCCTTGATCCCTCTCTGATT 368

190 190

369 CAGGGGAAAGGAACCTCCGAGTGAAGGAAGGTAGATTGGGAAAAGG 418

191GluThrThrTyrTrpLys.....ProProlLysAsnLysT 202

419 GAANGATTCCACTTGACAGAAAGTGGGACAGACTCTCTCCAAGAGTAGAG 468

202 rPTpAspGly 205

469 CTGGAAGGGA 479

seq_name: gb_est21:AA936814

seq_documentation_block:

LOCUS AA936814 405 bp mRNA EST 19-MAY-1998
DEFINITION oh89c05.s1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1474184 3', similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN (HUMAN);, mRNA

SEQUENCE

ACCESSION AA936814

NID 93094848

VERSION AA936814.1 GI:3094848

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 405)

AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT On Jan 17, 1998 this sequence version replaced gi:2045665.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www.bio.llnl.gov/dbrrp/image/image.html

Insert Length: 990 Std Error: 0.00

Seq primer: -40ml3 fwd. Et from Amersham

High quality sequence stop: 378.

Location/Qualifiers

1. .405

/organism="Homo sapiens"

/db_xref="taxon:9606"

/map="3: 9"

/clone="IMAGE:1474184"

/clone_lib="NCI_CGAP_Co8"

/tissue_type="adenocarcinoma"

/lab_host="DH10B"

/note="Organ: Colon; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from

colon adenocarcinoma, and was then primed with a Not I -

oligo(dT) primer. Double-stranded cDNA was ligated to Eco

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40UP from Gibco.
 Location/Qualifiers
 1..295

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 408)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierrey-Mieg, J.,
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
 and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478

COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1800
 Email: est@watson.wustl.edu
 Insert Size: 32
 High quality sequence stops: 297 Source: IMAGE Consortium, LLNL This
 clone is available royalty-free through LLNL; contact the IMAGE
 Consortium (info@image.llnl.gov) for further information.
 Seq primer: -21ml3
 High quality sequence stop: 297.
 Location/Qualifiers
 1..408
 /organism="Homo sapiens"
 /db_xref="GDB:501748"
 /db_xref="taxon:9606"
 /clone="IMAGE:84691"
 /clone_lib="Stratagene liver (#937224)"
 /sex="male"
 /dev_stage="49 years old"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: liver; Vector: pBluescript SK; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dr. Hepatectomy from normal male caucasian. Average
 insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 99 a 48 c 75 g 73 t
 ORIGIN

alignment_scores:
 Quality: 83.50 Length: 98
 Ratio: 1.856 Gaps: 4
 Percent Similarity: 45.918 Percent Identity: 29.592

alignment_block:

US-09-209-961-3 x AI589903 ..

Align seg 1/1 to: AI589903 from: 1 to: 295

75 TptTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87

18 TGGCTCTTGGGG...CGATGTCATCTTTTCATATGAAAGAACAGCAAA 64

87 87

65 GTTCAACACAAATAGAAATCTCAATGTAGTAGAACAAACCAAGTG 114

88GlyThrAspGlnAsnLysGlu.....TyrCysSerLys 99

115 TGTGAGGGGGGAACACACAAAGGAAGAAATGAGATGTTCGCAAAAA 164

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116

165 GATGGAGGA.....GGTTCCCTCTCTCTGCGGGACTGA 199

116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133

200 CTCAACACATGTGTGCGCATATACACCATCCAGAGTCAGGGGTGTCA 249

133 alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147

250 TTCTTTTGGGAGTAGAAAAAGGTGGGGATTAAAGAGCGTTT 293

seq_name: gb_est2:T74565

seq_documentation_block:

LOCUS T74565 408 bp mRNA EST 02-MAR-1995
 DEFINITION yc56d10.s1 Stratagene liver (#937224) Homo sapiens cDNA clone
 IMAGE:84691 3', similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN
 (HUMAN);, mRNA sequence.

ACCESSION T74565

NID 9691240

VERSION T74565.1 GI:691240

KEYWORDS EST.

SOURCE human.
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 408)
 AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierrey-Mieg, J.,
 Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.
 and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
 JOURNAL Genome Res. 6 (9), 807-828 (1996)
 MEDLINE 97044478

COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1800
 Email: est@watson.wustl.edu
 Insert Size: 32
 High quality sequence stops: 297 Source: IMAGE Consortium, LLNL This
 clone is available royalty-free through LLNL; contact the IMAGE
 Consortium (info@image.llnl.gov) for further information.
 Seq primer: -21ml3
 High quality sequence stop: 297.
 Location/Qualifiers
 1..408
 /organism="Homo sapiens"
 /db_xref="GDB:501748"
 /db_xref="taxon:9606"
 /clone="IMAGE:84691"
 /clone_lib="Stratagene liver (#937224)"
 /sex="male"
 /dev_stage="49 years old"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: liver; Vector: pBluescript SK; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dr. Hepatectomy from normal male caucasian. Average
 insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
 sequence: 5' GAATTCGACGAG 3' -3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3'"

BASE COUNT 118 a 71 c 109 g 99 t 11 others
 ORIGIN

alignment_scores:
 Quality: 85.00 Length: 100
 Ratio: 1.848 Gaps: 3
 Percent Similarity: 46.000 Percent Identity: 28.000

alignment_block:

US-09-209-961-3 x T74565 ..

Align seg 1/1 to: T74565 from: 1 to: 408

75 TptTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87

3 TTTTGGCTCTTNGGGCGATGTCATCTTTTCAATATGAAAAAGACGCAAA 52

87 87

53 GTTCAACATAAAATAGAAATCTCAATNTAGTAGAACAAACCAAGTT 102

88GlyThrAspGlnAsnLysGlu.....TyrCysSerLys 99

103 TNTGAGGGGGGAACACACAGCAAGGAAGAAATGAGATGTTCGCAAAAA 152

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116

153 GATGGAGGA.....GGTTCCCTCTCTCTGCGGGACTGA 187


```

FEATURES
  source      Location/Qualifiers
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR48N07"
/note="end: 17"
BASE COUNT  260 a 288 c 258 g 280 t 15 others
ORIGIN
alignment_scores:
  Quality: 89.50      Length: 301
  Ratio: 0.683        Gaps: 15
  Percent Similarity: 43.522  Percent Identity: 22.259
alignment_block:
US-09-209-961-3 x CNS00Ltz/rev ..
Align seg 1/1 to reverse of: CNS00Ltz from: 1 to: 1101
1 MetProSerLysLysAsnGlyArgSerGlyProGlnProHisLysArgTr 17
  :: ::::::::::::::::::::
806 CTGAACAGAGAGAGATGATGAGAGTATACCGGAGAGAGATCGC..... 762
17 pValPheThrLeuAsnAsnProSerGluAspGluArgLysLysLysLea 34
  ::::::::::::::::::::
761 .....TTCGCCACCGTCGATGAAAGAGAAATGTTAGAGG 728
34 luLeuProIleSerLeuPheAspTyrPheIleValGlyGluGluGlyAsn 50
  ::::::::::::::::::::
727 AGCTG...ATCAATATGATCATCAATAGTATTATTAAAGGAACCTCCGGA 681
51 GluGluGlyArgThrProHisLeuGlnGlnGlnPheAlaAsnPheValLys 67
  :: ::::::::::::::::::::
680 CCACACCCCAAGACACACATCTCTATAAGCCGCCCAAAAGGTCCTCCT 631
67 sGlnThrPheAsnLysValLysTyrPheLeuGlyAlaArgCysHisIleG 84
  ::
630 CGGC.....
84 luLysAlaLysGlyThrAspGlnGlnAsnLysGluTyrCysSerLysGlu 100
  ::::::::::::::::::::
626 .....TCGGATCAAGAGAT.....CTGAGGCT 603
101 GlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnArgSe 117
  ::::::::::::::::::::
602 CTGGCGCTAAGGATCAAGCGGGCTATGAGCTCGCGCAGGAGGAGGAGA 553
117 rAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuValt 134
  ::::::::::::::::::::
552 GGATCTGAAGAGA.....GTGTAA 533
134 hrValAlaGluGlnHisProValThrPheValLysAsnPheArgGlyLeu 150
  ::::::::::::::::::::
532 CGGTGGAGGAGGAG.....GTGGCTCTCAAGACCACTTAAGCGTTA 489
151 AlaGluLeuLeuLysValSerGlyLysMetGlnLysArgAspTrpLysTh 167
  ::::::::::::::::::::
488 CCGGAC.....GGCTCCTGG..... 474
167 rAsnValHisPheIleValGly.ProProGlyCysGlyLysSerLysTrp 183
  ::::::::::::::::::::
473 .....ATCATTCTGGGAACACCCC..... 455
184 AlaAlaAsnPheAlaAsnProGluThrThrTyrTrpLysProLysAs 200
  ::::::::::::::::::::
454 .....AACTTC.....CCCACTTGGAC 438
200 nLysTrpTrpAspGlyTyrHisGlyGluLysValValIleAspAsp 217
  ::::::::::::::::::::
437 GAGGAGATGGCCAAACTGGGAG.....GACTCCT 409

```

```

217 heTyrGlyTrpLeuProTrpAspAspLeuLeuArgLeuCysAspArgTyr 233
  :: ::::::::::::::::::::
408 GGCTGCTTGGTAGCCATGG.....CTGTTT 383
234 ProLeuThrValLysThrLysGlyGlyThrValProPheLeuAlaArgSe 250
  ::::::::::::::::::::
382 CAGCAAGATATAAGCTGTCCACCTGGA.....ATTCTGAACATAATC 342
250 rIleLeuIleThrSerAsnGlnThrProLeuGluTrpTyrSerSerThrA 267
  ::::::::::::::::::::
341 GAGTTTGGTGGAACTGCAACAGCTTTGGAGAAGATTACCAAACTA 292
267 laValProAlaValGluAlaLeuTyrArgArg.ileThrSerLeuValPh 283
  ::::::::::::::::::::
291 CGAGTACATATTGCTTGCCTATTGGCCAGCAAGATTACCATATATGCCA 242
283 eTrpLysAsnAlaThrLysGlnSerThrGluGluGlyGlyGlnPheVal 299
  ::::::::::::::::::::
241 TTGGCGAGAGCAACTTCAGTTTGGAGGTGAAGAGGTGGTCAATCCTC 193
seq_name: gb_est11:AA226908
seq_documentation_block:
LOCUS      AA226908          372 bp      mRNA          EST      11-MAR-1998
DEFINITION zrl9e04.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens
            cDNA clone IMAGE:663870 3' similar to gb:J00314_rna2 TUBULIN BETA-1
            CHAIN (HUMAN); mRNA sequence.
ACCESSION  AA226908
NID        91848444
VERSION    AA226908.1 GI:1848444
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 372)
AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisels,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.
            WashU-NCI human EST Project
            Unpublished (1997)
            On Sep 12, 1996 this sequence version replaced gi:1404680.
Contact:  Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2293 Std Error: 0.00
Seq primer: -41m13 fwd. Et from Amersham.
FEATURES
  source      Location/Qualifiers
1..372
/organism="Homo sapiens"
/db_xref="GDB:5425681"
/db_xref="taxon:9606"
/clone="IMAGE:663870"
/tissue_type="neuroepithelial cells"
/dev_stage="Ntera-2 neuroepithelial cells"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: brain; Vector: pBluescript SK-; Site: 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Uninduced, exponentially growing neuroepithelial
cells (Ntera-2/cl.D1). Average insert size: 1.0 kb;
Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGACGAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
BASE COUNT  109 a 74 c 91 g 98 t
ORIGIN

```



```

alignment_scores:
  Quality: 83.50      Length: 98
  Ratio: 1.856        Gaps: 4
  Percent Similarity: 45.918  Percent Identity: 29.592

alignment_block:
US-09-209-961-3 x AA226908 ..

Align seg 1/1 to: AA226908 from: 1 to: 372

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
25 TGGCTCTTGGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAA 71
87 .....

72 GTTCAACACAAATAGAAATCTCAATGTAGGATAGAACAAACCAAGTG 121
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
122 TGTGAGGGGGGAGCAGCAGCAAGAAAGGAGAAATGAGATGTTGCAAAAAA 171

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
172 GATGAGGAG.....GGTTCCCTCTCTCTCTGGGGACTGA 206

116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
||||| :|||:||||| ||||| ||||| ||||| ||||| |||||
207 CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTTC 256

133 alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
257 TCTCTTTTGGGAGTAGAAAAGGTGGGATTAAAGACGTTT 300

seq_name: gb_est7:W20432

```

```

seq_documentation_block:
LOCUS      W20432      379 bp      mRNA      EST      03-MAY-1996
DEFINITION z668005.s1 Soares_fetal_heart_NDHH19W Homo sapiens cDNA clone
IMAGE:327464 3' similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN
(HUMAN);, mRNA sequence.

ACCESSION  W20432
NID        g1295049
VERSION    W20432.1 GI:1295049
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 379)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
            Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
            Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
            Trevaskis,E., Waterston,R., Williamson,A., Wohlmann,P. and
            Wilson,R.
TITLE     The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   On Nov 29, 1993 this sequence version replaced gi:430432.

```

```

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: mob.REGA+ET
High quality sequence stop: 325.
Location/Qualifiers
1..379
/organism="Homo sapiens"

```

```

/db_xref="GDB:126166"
/db_xref="taxon:9606"
/clone="IMAGE:327464"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a
modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTACCAATCTGAAGTGGAGCGCCGATCTTTTCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT7T3 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT  113 a 76 c 96 g 94 t
ORIGIN

```

```

alignment_scores:
  Quality: 83.50      Length: 98
  Ratio: 1.856        Gaps: 4
  Percent Similarity: 45.918  Percent Identity: 29.592

alignment_block:
US-09-209-961-3 x W20432 ..

Align seg 1/1 to: W20432 from: 1 to: 379

75 TrpTyrLeuGlyAlaArgCysHisIleGluLysAlaLys..... 87
||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
18 TGGCTCTTGGGG...CGATGTCATCTTTTCAATATGAAAAAAGCAGCAA 64
87 .....

65 GTTCAACACAAATAGAAATCTCAATGTAGGATAGAACAAACCAAGTG 114
88 .....GlyThrAspGlnGlnAsnLysGlu.....TyrCysSerLys 99
||||| ||||||| ||||| ||||| ||||| ||||| ||||| |||||
115 TGTGAGGGGGGAGCAGCAGCAAGAAAGGAGAAATGAGATGTTGCAAAAA 164

100 GluGlyAsnLeuLeuIleGluCysGlyAlaProArgSerGlnGlyGlnAr 116
:::||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
165 GATGAGGAG.....GGTTCCCTCTCTCTCTGGGGACTGA 199

116 gSerAspLeuSerThrAlaValSerThrLeuLeuGluSerGlyIleLeuV 133
||||| :|||:||||| ||||| ||||| ||||| ||||| |||||
200 CTCAAACACTGATGTGGCAGTATACACCATTCAGAGTCAGGGGTGTTC 249

133 alThrValAlaGluGlnHisProValThrPheValLysAsnPhe 147
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
250 TCTCTTTTGGGAGTAGAAAAGGTGGGATTAAAGACGTTT 293

seq_name: gb_gss7:AQ469737

```

```

seq_documentation_block:
LOCUS      AQ469737      505 bp      DNA      GSS      22-APR-1999
DEFINITION CITBI-EL-2584G10.TR CITBI-El Homo sapiens genomic clone 2584G10,
genomic survey sequence.
ACCESSION  AQ469737
NID        g4653391
VERSION    AQ469737.1 GI:4653391
KEYWORDS   GSS.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE  1 (bases 1 to 505)
AUTHORS   Zhao,S., Adams,M.D., Nierman,W., Malek,J., Shizuya,H., Simon,M. and

```

Venter, J.C.
Use of BAC End Sequences from Caltech Libraries for Sequence-Ready
Map Building
Unpublished (1997)
Other_GSSs: C17B1-E1-2584G10.TF
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbeetigr.org
Clones are available from Research Genetics (info@resgen.com). BAC
end search page:
http://www.tigr.org/tldb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M3 Reverse
Class: BAC ends.

FEATURES

Location/Qualifiers
1..505
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2584G10"
/clone_lib="C17B1-E1"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelOBAC11; Site_1: ECORI; Site_2: ECORI;
Caltech Human BAC Library D"

BASE COUNT 75 a 148 c 135 g 147 t

alignment_scores:

Quality: 85.00 Length: 109
Ratio: 1.371 Gaps: 5
Percent Similarity: 56.881 Percent Identity: 28.440

alignment_block:

US-09-209-961-3 x A0469737/rev ..

Align seg 1/1 to reverse of: A0469737 from: 1 to: 505

```
11 ProGlnProHisLysArgTrpValPheThrLeuAsnAsnProSerGluAs 27
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
375 CCCCTCCCTACAGACAGTGGCTGCTCCATCTCCTCCTGCTCTCGG 326
27 pGluArgLysLysLeuArgGluLeuProIleSerLeuPheAspTyrPhe 44
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 CAGCCGACGACGAGAGTGGG...TGTGCAGCAGCAATTTGGAACACAGATGG. 280
44 leValGlyGluGluGluGluGluGluGluGluGluGluGluGluGlu 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
279 .....CAGAGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 238
61 PheAlaAsnPheValLysLysGlnThrPheAsnLysVal...LysTrpTyr 76
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
237 .....GAGACACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGATG 206
76 rLeuGlyAlaArgCysHisIleGluLysAlaLysGlyThrAspGlnGln 93
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
205 GCAGGGCCCAACATGTCTGCTGAGCCCCCAACCCAGGGGGCAAGACA 156
93 snLysGluTyrCysSerLysGluGluGluGluGluGluGluGluGlu 109
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 GTAAGATGAGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGT 115
110 ProArgSerGlnGlyGlnArgSerAsp 118
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 AAACGGATGCAAGCCCAAGCAAGGAA 88
```

seq_name: gb_est11:AA218011

seq_documentation_block: 667 bp mRNA EST 06-FEB-1997
LOCUS AA218011

DEFINITION

mv59e06.r1 Soares mouse 3NME12 5 Mus musculus cDNA clone
IMAGE:659362 5', mRNA sequence.

ACCESSION AA218011
NID 91827020
VERSION AA218011.1 GI:1827020
KEYWORDS EST.
SOURCE house mouse.

ORGANISM

Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE

1 (bases 1 to 667)
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisler, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1393071.

TITLE

WashU-HMI Mouse EST Project

JOURNAL

Unpublished (1996)

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1393071.

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MG1:405210

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 494.

FEATURES

source

1..667
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:659362"
/clone_lib="Soares mouse 3NME12 5"
/sex="unknown"
/tissue_type="fetus"
/dev_stage="12.5dpc total fetus"
/lab_host="DH108"
/note="Organ: whole fetus; Vector: pT73D-pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5' TGTTACCAATCTGAAGTGGGCGGCGCTATTTTCTTTTCTTTT
3'], on total mouse RNA [provided by Minoru Ko, Wayne
State Univ.]; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
Library went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaudo."

BASE COUNT 151 a 168 c 208 g 140 t

ORIGIN

alignment_scores:

Quality: 86.50 Length: 177
Ratio: 0.951 Gaps: 8
Percent Similarity: 51.412 Percent Identity: 26.554

alignment_block:

US-09-209-961-3 x AA218011 ..

Align seg 1/1 to: AA218011 from: 1 to: 667

```
66 LysLysGlnThrPheAsnLysValLysTrpTyrLeuGlyAlaArgCysHi 82
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
124 CGGGAGGAGGAGTGAATGCTCTGCTGGGCGCTGGGCGGAG.... 168
82 sileGluLysAla.LysGlyThrAspGlnGlnAsnLysGluTyrCys... 97
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```

seq_name: gb_est5:H95640

seq_documentation_block:
  LOCUS       H95640          389 bp      mRNA          25-NOV-1996
  DEFINITION   Yv17a07.s1 Soares fetal liver spleen 1NF1S Homo sapiens cDNA clone
  IMAGE:242964 3' similar to gb:J00314_rna2 TUBULIN BETA-1 CHAIN
              (HUMAN);, mRNA sequence.
  ACCESSION   H95640
  NID         g1108782
  VERSION     H95640.1  GI:1108782
  KEYWORDS    EST.
  SOURCE      human.
  ORGANISM    Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Chordata; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 389)
  Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
  Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
  Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
  Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
  Wilson, R.
  TITLE       The WashU-Merck EST Project
  JOURNAL     Unpublished (1995)
  COMMENT     On Apr 14, 1993 this sequence version replaced gi:635738.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
High quality sequence stops: 257
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

```

OM of: US-09-209-961-5 to: GenEmbl:* out_format : pfs

Date: Dec 27, 1999 2:13 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:

-MODEL=frame+2n.model -DEV=xlp
-Q=/cgn2.1/USPTO_spool/US09209961/runat_22121999_101043_17910/app_query.fasta.1
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -YGAPEXT=0.500 -FGAPOP=6.000
-FGAPOP=7.000 -XGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOALIGN=200 -THR_SCORE=escore -ALIGN=15 -MODE=LOCAL
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-5

Query length: 233

Database: GenEmbl.*

Database sequences: 780561

Database length: 2137953050

Search time (sec): 1994.760000

score_list:

Sequence	Strd Orig	Zscore	EScore	Len	Documentation
gb_vi:AF085695	- 1280.00	1907.73	2.9e-98	1768	AF085695 Porcine circovirus st
gb_vi:AF086836	- 1280.00	1907.73	2.9e-98	1768	AF086836 Porcine circovirus st
gb_vi:AF086835	- 1276.00	1901.73	6.3e-98	1768	AF086835 Porcine circovirus st
gb_vi:AF086834	- 1257.00	1873.23	2.4e-96	1768	AF086834 Porcine circovirus st
gb_vi:PCJ3185	- 1251.00	1864.23	7.7e-96	1768	AF223185 Porcine circovirus st
gb_vi:AF166528	- 1244.00	1853.73	3.0e-95	1768	AF166528 Porcine circovirus st
gb_vi:AF027217	- 1237.00	1843.23	1.1e-94	1768	AF027217 Porcine circovirus st
gb_vi:AF112862	- 1237.00	1843.23	1.1e-94	1768	AF112862 Porcine circovirus st
gb_vi:AF147751	- 1234.00	1838.73	2.0e-94	1768	AF147751 Porcine circovirus st
gb_vi:AF109397	- 1230.00	1832.73	4.4e-94	1768	AF109397 Bovine circovirus, c
gb_vi:AF154679	- 1226.00	1826.73	9.5e-94	1768	AF154679 Porcine circovirus, c
gb_vi:AF109399	- 1217.00	1813.23	5.4e-93	1768	AF109399 Porcine circovirus, c
gb_vi:AF109398	- 1207.00	1798.23	3.7e-92	1768	AF109398 Porcine circovirus ty
gb_vi:AF117753	- 1202.00	1790.73	9.6e-92	1768	AF117753 Porcine circovirus ty
gb_vi:AF118097	- 1163.00	1740.71	5.9e-89	700	AF118097 Porcine circovirus typ
gb_vi:AF012107	- 848.50	1260.51	3.3e-62	1759	AF012107 Porcine circovirus, c
gb_vi:PCCOMGEN	- 834.50	1239.51	4.9e-61	1759	Y09921 Porcine circovirus comp
gb_vi:PCU49186	- 833.50	1238.01	5.9e-61	1759	U49186 Porcine circovirus, com
gb_vi:AF071879	- 820.50	1218.52	7.2e-60	1758	AF071879 Porcine circovirus, c
gb_vi:AF055392	- 692.00	1025.71	3.9e-49	1768	AF055392 Porcine circovirus ty
gb_vi:AF055391	- 681.00	1009.21	3.3e-48	1768	AF055391 Porcine circovirus ty
gb_vi:AF055393	- 669.00	991.21	3.3e-47	1767	AF055393 Porcine circovirus Ty
gb_vi:AF055394	- 669.00	991.21	3.3e-47	1767	AF055394 Porcine circovirus Ty
gb_vi:AF135393	- 398.00	597.48	2.8e-25	438	AF135393 Porcine circovirus PCV
gb_vi:AF134203	- 365.00	548.02	1.6e-22	436	AF134203 Porcine circovirus PCV
gb_vi:AF071878	- 226.50	326.34	3.5e-10	1993	AF071878 Beak and feather dise
gb_vi:AF080560	- 214.00	307.59	3.9e-09	1993	AF080560 Beak and feather dise
gb_ba:SCU51728	+ 96.50	133.65	19.20	1549	U51728 streptomyces coelicolor
gb_vi:BYV223589	+ 89.50	131.79	24.37	603	AJ223589 Barley yellow dwarf vi
gb_p12:CN501525	+ 88.50	130.79	27.71	571	AF001525 Musa acuminata ripenin
gb_p12:CN5018BA	+ 87.50	130.21	29.82	516	AL14230 Botrytis cinerea stral
gb_vi:BYDMVCPG	+ 87.00	128.08	39.20	600	X17260 Barley yellow dwarf viru
gb_vi:BYV223588	+ 86.50	127.29	43.41	603	AJ223588 Barley yellow dwarf vi
gb_vi:BYV223587	+ 85.50	125.79	52.62	603	M21347 Barley yellow dwarf viru
gb_ba:BYDPTCT	+ 84.50	125.40	55.30	534	U13939 Mycobacterium tuberculo
gb_ba:MTU13939	+ 84.50	125.40	55.30	534	U13939 Mycobacterium tuberculo
gb_pai:120874	+ 91.50	124.53	61.82	1848	I20874 Sequence 1 from patent
gb_pai:BACASG	+ 94.00	124.40	62.88	2824	D38058 Bacillus caldotenax gen
gb_vi:BYDPCPG	+ 87.50	124.30	63.66	984	X56050 Barley yellow dwarf viru
gb_vi:BYD7918	+ 84.50	124.29	63.78	603	AJ007918 Barley yellow dwarf vi
gb_vi:BYD7929	+ 84.50	124.29	63.78	603	AJ007929 Barley yellow dwarf vi
gb_p11:ZMPGAL2	+ 90.00	123.82	67.71	1562	X57628 z.mays mRNA for polygal
gb_vi:AF141385	+ 84.00	123.72	68.58	591	AF141385 Sugarcane yellow leaf

gb_pat:109604 + 91.50 123.70 68.79 2024 ! I09604 Sequence 4 from Pate
gb_p11:ANGOX + 91.50 123.70 68.79 2024 ! X16061 Aspergillus niger go
gb_om:CFA388555 + 92.00 123.61 69.57 2218 ! AJ388555 Canis familiaris m

seq_name: gb_vi:AF085695

seq_documentation_block:
LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998
DEFINITION Porcine circovirus strain 412, complete genome.
ACCESSION AF085695

NID 93668362

VERSION AF085695.1 GI:3668362

KEYWORDS

SOURCE porcine circovirus.

ORGANISM porcine circovirus

REFERENCE 1 (bases 1 to 1768) Viruses; ssDNA viruses; Circoviridae; Circovirus.

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,

University of Saskatchewan, 120 Veterinary Road, Saskatoon,

Saskatchewan S7N 5E3, Canada

FEATURES Location/Qualifiers

source

1..1768 /organism="porcine circovirus"

/strain="412"

/db_xref="taxon:46221"

51..995 /codon_start=1

/product="P35.8"

/protein_id="AAC61860.1"

/db_xref="PID:g3668363"

/db_xref="GI:3668363"

/translation="MPKNGRSQPHKPVFTLNPSDEKKIRELPISLDFYFI

VGEENEGRTPHLQGFANFVKKOTFNKVKYLGARCHIEKAKGTQDNKKYCSKEGN

LLIECGAPRSGQSDLSSTLESGLTVAKQHPVTVKFNKRLAEALLKYSVSGKM

LBKCKNTYHFTVGGPCGSKWAANFANPETTYKPPKNKWDGYSKGVVDDFY

GWLPDRYDLRLCDRPLTVTKGTGTPFLARSILTSSTAVPAVEALRY

RITSLVFWKNAKSTEGGQFVLSPPCPPEFYNY"

complement(357..671)

/codon_start=1

/product="P12.1"

/protein_id="AAC61865.1"

/db_xref="PID:g3668368"

/db_xref="GI:3668368"

/translation="MVTPPLVFRWFPVCGFRVCKISSPFAFTTPRPHNEVYIGFPI

TLHFPAHFQKFSQPAIEFDKRYRVLICNGHQNALQOQTHSSHQVPLSLRSRSTF

NK"

complement(385..565)

/codon_start=1

/product="P6.6"

/protein_id="AAC61866.1"

/db_xref="PID:g3668369"

/db_xref="GI:3668369"

/translation="MKCTLVFSQSFIFPLTFKSSAPRKELTNVTCGFATVTRIPL

SNKVLTAVDRLRCP"

553..732 /codon_start=1

/product="P6.6"

/protein_id="AAC61861.1"

/db_xref="PID:g3668364"

/db_xref="GI:3668364"

/translation="MTYSLMHLGVVRANGLLLIQTAKPHGNHLKTSGGMYTVKWKW

LLMTFFMAGRCGMV"

833..970 /codon_start=1

/product="P5.0"

/protein_id="AAC61862.1"

/db_xref="PID:g3668365"

/db_xref="GI:3668365"
/translation="NVLLNCCPSCRSLSLSEDFYFLGLECYKTIHGGRGVRRPFP
P"

CDS 1016..1177

/codon_start=1
/product="p6.2"
/protein_id="AAC61863.1"
/db_xref="PID:g3668366"
/db_xref="GI:3668366"
/translation="MVFIHLGFKWGVKIFSELYIHGYTDIVLVVFTVFS
SAEAYVHISRL"

CDS complement(1034..1735)

/codon_start=1
/product="p27.9"
/protein_id="AAC61864.1"
/db_xref="PID:g3668367"
/db_xref="GI:3668367"
/translation="MTYPRRRYRRRRHRRSHLQILRRRPLVHPRHRYR
WRKNGI
ENTRLSRTFGYTVKRTTTPSWAVDMRFKIDDFVPGGGTNKISIP
EYVIRKVK
VEFWPCSPITQDRCGVGSTAVILDDNEVTKATALYDVPYNTSSRHT
IQPFYSHSRY
FTPKVLSTIDYDFQPNKRNLWLRLQTSNVDHVLGTFATFENSKID
QDYNIRVTMI
VOFREFNKDPPLP"

BASE COUNT 463 a 362 c 481 g 462 t
ORIGIN

alignment_scores:
Quality: 1280.00 Length: 233
Ratio: 5.494 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-5 x AF085695/rev ..

Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

1 MetThrProArgArgArgTyrArgArgArgHisArgProArgSe 17
1735 ATACGCTATCCAGGAGCGGTACCGCAGAAAGACACCCCGCCGAG 1686
17 rHisLeuGlyGlnLeuArgArgProTrpLeuValHisProArgH 34
1685 CCATCTGGCAGATCTCGCGCGCGCTGCTGCTCCACCCCGCC 1636
34 isArgTyrArgTrpArgGlyAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGCTACCTGGAGAAAGAAATGGCATCTTCAACACCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrTrpSerTr 67
1585 CGCACCTTCGGATATAGTCAAGCGTACCACATCAACGCCCTCTCG 1536
67 pAlaValAspMetMetArgPheLysIleAspPheValProGlyG 84
1535 GCGGCTGGACATGATGAGATTAAATGACGACTTGTCTCCCGGAG 1486
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAAATCTCTATACCTTGAATCTACAGAAATAAGAA 1436
101 valLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAAGTTGAATCTCGCTTGTCTCCCATCACCGGCTGATAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspPheValThrLysAla 134
1385 AGTGGCTCCACTGCTGTTATTTAGATGATACTTGTAAACAAGGCCA 1336
134 hrAlaLeuThrTrpAspProTyrValAsnTyrSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCATATGTAAACTACTCTCCCGCCATACATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProVal 167

1285 CCCAACCCCTCTCTACCACTCCCGTTACTTACACCCAAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGATCCCACTATTGATTACTTCCACCAAAATAACAAAAGGAATCAGCTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GCGTAGGCTACAACCTCTGGAATGTGGACACGCTAGGCTCGGCAT 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 CGTTCCGAAACAGTAATACGACGAGCTACAATATCCGTGTAAACAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACAATTACAGAAATTAATCTTAAAGACCCCACTTGAACCC 1037
seq_name: gb_vi:AF086836

seq_documentation_block:

LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain M226, complete genome.
ACCESSION AF086836
NID 93661521
VERSION AF086836.1 GI:3661521
SOURCE porcine circovirus.
ORGANISM Porcine circovirus
REFERENCE 1 (bases 1 to 1768)
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES
Source

1..1768
/organism="porcine circovirus"

/strain="M226"

/db_xref="taxon:46221"

51..995

/codon_start=1

/product="p35.8"

/protein_id="AAC61741.1"

/db_xref="GI:3661522"

/db_xref="GI:3661522"

/translation="MPSKNGRSGQPQHRKRVFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRTPHLOGFANFVKQTKVYFGARCHEKAGTKDQONKEYSKRGK
LLECGAPRSQGSRLSTAVSTLSEGLVTVAKOHPVTEVKNRGLAKLLKVSCKM
QRDKWTNVHFIIVGPPCGCKSKWAANFANPETYKPKPKKWDGSGYGEKVVYIDFY
GWLDDLLRLCDRLPLTVTKTGTVFFLARSILITSNQTPLEWISSTAVPAVEALYR
RITSLVFWKNATQSTEGGQFVTLSPPCPEPPEYIN"

complement(1034..1735)

/codon_start=1

/product="p27.9"

/protein_id="AAC61742.1"

/db_xref="PID:g3661523"

/db_xref="GI:3661523"

/translation="MTYPRRRYRRRRHRRSHLQILRRRPLVHPRHRYRWRKNGI
ENTRLSRTFGYTVKRTTTPSWAVDMRFKIDDFVPGGGTNKISIPPEYVIRKVK
VEFWPCSPITQDRCGVGSTAVILDDNEVTKATALYDVPYNTSSRHTIQPFYSHSRY
FTPKVLSTIDYDFQPNKRNLWLRLQTSNVDHVLGTFATFENSKIDQDYNIRVTMI
VOFREFNKDPPLP"

BASE COUNT 463 a 362 c 480 g 462 t
ORIGIN

alignment_scores:

Quality: 1280.00 Length: 233
Ratio: 5.494 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-5 x AF086836/rev ..

Align seg 1/1 to reverse of: AF086836 from: 1 to: 1768

```
1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
|||||
1735 ATGACGTATCCAAAGGAGCGTTACCGCAGAAGACACACGCCCGCCGAG 1686
|||||
17 rHisLeuGlyGlnIleLeuArgArgArgProTrrLeuValHisProArgH 34
|||||
1685 CCATCTTGGCCAGATCTCCGCGCGCGCCCTGGCTGCTCCACCCGCC 1636
|||||
34 isArgTyrArgTrrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGCTACCGTTGGAGAGGAAAAATGGCATCTTCAACACCCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585 CGCAGCTTCGGATATCTGTCAGCGTACACAGTACACACGCCGCTCTG 1536
|||||
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
|||||
1535 GCGGTGGACATGATGAGATTAAATTGACGACTTGTTCGCCCGGAG 1486
|||||
84 lYgLyThrAsnLysIleSerIleProPheGlyTyrTrrArgIleArgLys 100
|||||
1485 GGGGACCAACAAAATCTCTATACCTTTGAATACTACAGAAATAGAAAG 1436
|||||
101 ValLysValGluPheTrrProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGGTTGAATCTTCGGCTTGTCTCCCATCACCGAGGTGATAGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGGCTCCATCGCTGTTATTTTAGATGATAACTTTGTAAACAAAGCCA 1336
|||||
134 hrAlaLeuThrTrrAspProTrrValAsnTrrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCATATGTAATCTACTCTCCCGCCATACATC 1286
|||||
151 ProGlnProPheSerTrrHisSerArgTrrPheThrProLysProValLe 167
|||||
1285 CCCCAACCCCTTCTCCPACCATCCCGTTACTTCCACACCCAAACCTGT 1236
|||||
167 uAspSerThrIleAspTrrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGACTCCATATGTATGTTACTTCCAAACAAATACAAAAGGAATCAGCTT 1186
|||||
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GCGTGGGCTACAAACCTCTGGAATGTGACACCATGAGCCCTCGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTrrAspGlnAspTrrAsnIleArgValThrMe 217
|||||
1135 GCGTTCGAAAACAGTAATACGACGAGGACTACAATATCGGTGACCAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTCAGAGATTTAATCTTAAAGACCCCCCACTTGAACCC 1037
|||||
```

seq_name: gb_vi:AF086835

seq_documentation_block:

LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION AF086835
NID 93661518
VERSION AF086835.1 GI:3661518

KEYWORDS

SOURCE porcine circovirus.

ORGANISM

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Emergence of a new porcine circovirus

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Wang, L., Willson, P., Chow, B., Gibbons, E. and Babiuk, L.

TITLE Direct Submission

JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120

Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

FEATURES

Source

1..1768

/organism="porcine circovirus"

/strain="9741"

/db_xref="taxon:46221"

51..995

/codon_start=1

/product="P35.8"

/protein_id="AAC61739.1"

/db_xref="PID:93661519"

/db_xref="GI:3661519"

/translation="MPSKKNGRSGQPQPKRWVTLNPNSENKRKKIRELPISLFDYFI
VGEENEGRTPLQGFANFVKQTFENKVKVLGARCHIEKAKGTDOONKEYCSKEG
LLIEGAPRSQCQSDLSSTLLESGILVTVAKQHPVTFVKNFRGLAELLKVSCKM
QRDKWNTVHFIVGPPGCGKSKWAANFNPETIYKPKKNWDSYHGKVVVDDFY
GWLWDLLRLCDRYPPLVTKTKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSVFWKNATEQSTEGGQFVTLSPPCPEPPEYIN"

CDS

complement(1034..1735)

/codon_start=1

/product="P27.9"

/protein_id="AAC61740.1"

/db_xref="PID:93661520"

/db_xref="GI:3661520"

/translation="MTYPRRYRRRRHRRSHLGOILRRPRLVHPHRYRWRKNGI
FNTRLSRTFGTVKRTVTTPSWAVDMMRFKIDDFVPGCGTGNKISIPFYRIRKYK
VEFWPCSPITQDGRGVSATVILDDNFRTKALTALYDVPVNYSSRHTIPQPPSYHSRY
FTPKPVLDSITIDYFQPNKRNQLWLRLQTSNVDVHVLGTAFENSKYDQDYNIRVTM
VOFRENLKDPPLKP"

BASE COUNT

464 a 360 c 480 g 464 t

ORIGIN

alignment_scores:

Quality: 1276.00 Length: 233

Ratio: 5.476 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 99.571

alignment_block:

US-09-209-961-5 x AF086835/rev ..

Align seg 1/1 to reverse of: AF086835 from: 1 to: 1768

```
1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17
|||||
1735 ATGACGTATCCAAAGGAGCGTTACCGCAGAAGACACACGCCCGCCGAG 1686
|||||
17 rHisLeuGlyGlnIleLeuArgArgArgProTrrLeuValHisProArgH 34
|||||
1685 CCATCTTGGCCAGATCTCCGCGCGCGCCCTGGCTGCTCCACCCGCC 1636
|||||
34 isArgTyrArgTrrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGCTACCGTTGGAGAGGAAAAATGGCATCTTCAACACCCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585 CGCAGCTTCGGATATCTGTCAGCGTACACAGTACACACGCCGCTCTG 1536
|||||
67 pAlaValAspMetMetArgPheLysIleAspPheValProGlyG 84
|||||
```

```

1355 GCGGTGGACATGATGAGATTAAATGACGACTTGTTCCTCCCGGAG 1486
84  lvglyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCAACAAATCTCTATACCTTTGAATACTACAGATAAGAAAG 1436
101  ValLysValGluPheTyrProCysSerProIleThrGlnGlyAspArgG1 117
|||||
1435 GTTAAGTTGAATCTGCGCTGCTCCCATCACCAGGGGTAGAGGG 1386
117  yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
|||||
1385 AGTGGCTCCACTGCTGTTATTTAGATGATAACTTTGTAACAAAGGCCA 1336
134  hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCATATGTAACACTCTCCCTCCCGCCATACAATC 1286
151  ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
|||||
1285 CCCCAACCTTCTCTACACTCCCGTACTTACACCCAAACCTGTCT 1236
167  uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGACTCCACTATTGATTACTTCCACCAATAACAAAGGAATCAGCTTT 1186
184  rLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGAGCTACAAACCTCTGGAATGTGGACCAGTAGCGCTCGACAT 1136
201  AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 CGTTCGAAACAGTAATATACACGAGGACTACATATCCGTGTACCAT 1086
217  tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACATTCAGAGAATTAACTTTAAAGACCCCCCACTTAACCC 1037
seq_name: gb_v1:AF086834

seq_documentation_block:
LOCUS      AF086834      1768 bp      DNA      circular      VRL      29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.
ACCESSION  AF086834
NID        33661515
VERSION    AF086834.1  GI:3661515
KEYWORDS   porcine circovirus.
SOURCE     porcine circovirus.
ORGANISM   porcine circovirus
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE  1 (bases 1 to 1768)
AUTHORS   Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE     Emergence of a new porcine circovirus
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 1768)
AUTHORS   Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE     Direct Submission
SUBMITTED (26-AUG-1998) VIDO, University of Saskatchewan, 120
VETERINARY Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES   Location/Qualifiers
            source
            1..1768
            /organism="porcine circovirus"
            /strain="B9"
            /db_xref="taxon:46221"
            51..995
            /codon_start=1
            /product="P35.8"
            /protein_id="AAC61737.1"
            /db_xref="PID:g3661516"
            /db_xref="GI:3661516"
            /translation="MPSKKNRSGQPHKRWFTLNPNSEDERKKIRELPISLDFYI
            VGEFEEGRTPHLQGFANFVKQTFNKLWGLGARCHIEKAKGTDQONKEYCSKEGN
            LLIECAPRSQGRSDLSAVSLLESGLVTAVAKQHPVTFVKRGLAELLKVSGRM

```

```

OKRDWKNVHEIVGPPGCKSKWAANFANPETTYWKKPKKMWDCYHGKGVVDDFY
GMLPWDDLRLCDRYPLTVTKGTVPFLARSILITSNOTPLEWSSYTAVPAAVEALYR
RITSIVFWKNVTEQSTEGGQFVLSPPCPPEFYINY"
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:g3661517"
/db_xref="GI:3661517"
/translation="MTYPRRRYRRRRHPRSHLGOILRRRRLVHPRHRYRWRKNGI
FNRLSRFTGYTVKATVTTPSWADMMRFNIDFVPPGGTKNISIFEFYIRRRYK
VEFWPCSPITQDRCGVSTAVLDNFYTKATALTDPYVYNSRHTTTPQPFVSHSRY
FTPKPVLDSITDYFPNNKRNQLWLRLQTSRNVHDVGLGTAFENSKYDODYNIRVTY
VQFRENLKDPPLP"
BASE COUNT      463 a      357 c      482 g      466 t
ORIGIN
alignment_scores:
    Quality: 1257.00      Length: 233
    Ratio: 5.442      Gaps: 0
Percent Similarity: 99.142      Percent Identity: 98.283
alignment_block:
US-09-209-961-5 x AF086834/rev ..
Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768
1  MetThrTyrProArgArgTyrArgArgArgHisArgProArgse 17
|||||
1735  ATGACGTATCCAAAGAGGGCTTACCAGAGAAGACACCGCCCGCAG 1686
17  rHisLeuGlyGlnIleLeuArgArgProTyrLeuValHisProArgH 34
|||||
1685  CCATCTTGGCAGATCTCGCGCGCCCTGGCTCGTCACCCCGCCG 1636
34  isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635  ACCGCTACCGTTGGAGAAGAAAATGGCATCTTCAACACCCCGCTCTCC 1586
51  ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585  CGCACCTTCGGATATACTGTCAAGGCTACCACAGTCACACGCGCTCTG 1536
67  pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1535  GCGGTGGACATGATGAGATTAAATATGACGACTTTGTTCCTCCCGGAG 1486
84  lvglyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485  GGGGACCAACAAATCTCTATACCTTTGAATACTACAGATAAGAAAG 1436
101  ValLysValGluPheTyrProCysSerProIleThrGlnGlyAspArgG1 117
|||||
1435  GTTAAGTTGAATCTGCGCTGCTCCCATCACCAGGGGTAGAGGG 1386
117  yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
|||||
1385  AGTGGCTCCACTGCTGTTATTTAGATGATAACTTTGTAACAAAGGCCA 1336
134  hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335  CAGCCCTAACCTATGACCATATGTAACACTCTCCCTCCCGCCATACAATC 1286
151  ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
|||||
1285  CCCCAACCTTCTCTACACTCCCGTACTTACACCCAAACCTGTCT 1236
167  uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235  TGACTCCACTATTGATTACTTCCACCAATAACAAAGGAATCAGCTTT 1186
184  rLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200

```



```

1185 GCGTGGGCTACAAACCTCTAGAAATGTGCACACCTAGGCGCTCGGCAC 1136
|||||
201 AlapheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCGTTGCGAAACAGTAATACGACACGAGGACTACAAATATCGGTAAACCAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GATGTACAATTCAGAGAAATTAATCTTAAGACCCCGCCCACTTAACCC 1037
|||||

```

seq_name: gb_vi:PCAJ3185

```

seq_documentation_block:
LOCUS PCAJ3185 1768 bp DNA circular VRL 03-JUL-1998
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION AJ223185
NID g3293040
VERSION AJ223185.1 GI:3293040
KEYWORDS complete genome.
SOURCE porcine circovirus.
ORGANISM porcine circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Viruses; ssDNA viruses; Circoviridae; Circovirus.
TITLES Direct Submission
JOURNAL Submitted (02-JAN-1998) Morozov I., Veterinary Medical Research
Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
50011, U.S.A
REFERENCE 2 (bases 1 to 1768)
AUTHORS Morozov, I., Sirinarumit, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
Yoon, K.J. and Paul, P.S.
TITLES Detection of a novel strain of Porcine circovirus in pigs with
Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL Unpublished
FEATURES
Location/Qualifiers
source 1..1768
/organism="porcine circovirus"
/virus
/isolate="ISU-31"
/specific_host="Sus scrofa"
/db_xref="taxon:46221"
51..995
/note="ORF1"
/codon_start=1
/protein_id="CAA11157.1"
/db_xref="PID:e1310034"
/db_xref="PID:g3293041"
/db_xref="GI:3293041"
/translation="MPKKNGSGPQPKRWVFLNPNSEDERKKIRELPISLDFYFI
VGEENEGRTPHLQGFANFKVKTFFNFKVWYLGARHIEKAKGQDQNKKEYCSKEGN
LLIECGAPRSQGSRLSTAYSTLLESGSLVTAEPQHPVFNFRGLAEILLKVSKEGK
QKRWKTNVHVIVGPGCGSKWAANFADPTTYKPPRNKWDGYPHGEVYVVDIFY
GWLPPDLLRLCDRYPLTVETKGGTVPLARSILITSNQTPLEWYSSSTAVPAVEALYR
RISLVFNKNAETEGGQFVLSLPCPEPFYNI"
BASE COUNT 452 a 361 c 492 g 463 t
ORIGIN

```

```

alignment_scores:
Quality: 1251.00 Length: 233
Ratio: 5.439 Gaps: 0
Percent Similarity: 98.712 Percent Identity: 97.854

```

alignment_block:

US-09-209-961-5 x PCAJ3185/rev ..

Align seg 1/1 to reverse of: PCAJ3185 from: 1 to: 1768

```

1 MetThrTyrProArgArgTyrArgArgArgArgHisArgProArgSe 17
|||||
1735 ATGACGTATCCAGGAGGGGTTACCGCAGGAGACACCGCCCGCCGACG 1686

```

```

17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTTGGCCAGATCTCCGCGCGCCGCTGCTGCTGCACCCCGGCC 1636
|||||
34 isArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCCTACCGTTGGAGAGGAAAAATGCAATCTTCAACACCGCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585 CGCACCTTCGGATATCTGTCAGGCTACCACAGTCAGACGCGCTCTCTG 1536
|||||
67 palValAspMetMetArgPheLysIleAspAspPheValProProGlyG 84
|||||
1535 GCGCGTGGACATGATGAGATTATATGAGACTTTGTTCCCGGGAG 1486
|||||
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGAGCAACAAATCTCTATACCTTTGAATACTACAGAAATAAGAAAG 1436
|||||
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgGl 117
|||||
1435 GTTAAGTTGAATTTCTGCGCTGCTCCCATCACCAGGCGGTATAGGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
|||||
1385 AGTGGGCTCCACTCTCTTATCTAGATGATACTTTGTACAAAGGCCA 1336
|||||
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAAACCTATGACCATATGTAATACTACTCTCCCGCCATACAATC 1286
|||||
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
|||||
1285 CCCCAACCTTCTCTTACCCTCCCTTACTTCTACACCCCAACCTGTCT 1236
|||||
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGACTCCACTATTGATTACTTCCAAACCAATAACAAAGGAATACGCTTT 1186
|||||
184 rpLeuArgLeuGlnThrSertGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTAAGGCTACAAACCTCTAGAAATGTGGACACGCTAGGCGCTCGGCAC 1136
|||||
201 AlapheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCGTTGCGAAACAGTAATACGACGAGGACTACAAATATCGGTAAACCAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GATGTACAATTCAGAGAAATTAATCTTAAGACCCCGCCCACTTAACCC 1037
|||||

```

seq_name: gb_vi:AF166528

```

seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID g5739338
VERSION AF166528.1 GI:5739338
KEYWORDS porcine circovirus.
SOURCE porcine circovirus
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLES Complete nucleotide sequences of porcine circovirus Tainan strand
outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLES Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi

```

Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-Yi City
600, Taiwan

FEATURES

source Location/Qualifiers

1..1768
/organism="porcine circovirus"
/virion
/strain="Taiwan"
/db_xref="taxon:46221"
/country="Taiwan"
51..995

CDS

/codon_start=1
/product="putative Rep protein"
/protein_id="AA050432.1"
/db_xref="PID:g5739339"
/db_xref="GI:5739339"
/translation="MPSKNGRSGPQHKRWFTLNPNSEDERKKIRELPISLFDYFI
VGEEGNEGRTPHLOGFANFKVKTQFNKVFAGCHIEKAKGTDOONKEYCSKEGN
LLIECGAPRSGQSDLSSTAVSLLESGSLVTVAEQHPVTFVRNFRGLAELKVSQGM
OKRDKNTNHHVIGVPPCGSKKWAANFADPTTYWKPKNKWDGYHGEVVIIDDFY
GWLPPDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSSSTAVPAVEALYR
RITSLVFNKNAEQSTEGGQFVTLSPCPPEFYEINI"

BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN

alignment_scores:

Quality: 1244.00 Length: 233
Ratio: 5.385 Gaps: 0
Percent Similarity: 99.142 Percent Identity: 96.996

alignment_block:

US-09-209-961-5 x AF166528/rev ..

Align seg 1/1 to reverse of: AF166528 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
1735 ATGCGTATCCNAGGAGGGCTTTCGCGACGACGACCGCCCGCGAG 1686
17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
1685 CCATCTTGCCAGATCCTCCGCGCGCGCCCTGGCTCGTCGCCACCCCGCC 1636
34 IsArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGTTACCGCTGGAGAGAAATGGCATCTTCAACACCCGCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
1585 CGCACCTTCGGATACGTCAAGGCTCAACAGTCAGAACGCCCTCCGTG 1536
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
1535 GCGCGTGACATGATGAGATTATATTATTAACGACATTGTTCCCGCGGAG 1486
84 lyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAAATCTCTATACCTTTGAATCTACAGATAAGAAAG 1436
101 ValLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAGGTTGAATTCGCGCCCTGTTCCTCCCAATCACCGGGGTACAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTGGCTCCACCTGCTGTTATCTAGATGATAACTTTGTAACGACCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCGGTATGAACTACTCTCCCGCCCATACATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167

1285 CCCCAACCCCTTCTCTACCACTCCCGGTACTTTTACCCCAACCTGTCCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
1235 TGATCCACTATTGATTACTTCCACCAACAAACAAAGGAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GGTGAGGCTACAAACCTCGCAATGTGGACCACGTTAGCCCTGGGCAT 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTCGAACACAGTAATACGACGAGGACTACATATCGTGTAACTAT 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACATTCAGAGAATTTAATCTTAAGACCCCCCACTTAAACCC 1037
seq_name: gb_vi:AF027217
seq_documentation_block:
LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998
DEFINITION Porcine circovirus strain pmws PCV, complete genome.
ACCESSION AF027217
NID Q2689645
VERSION AF027217.1 GI:2689645
KEYWORDS porcine circovirus.
SOURCE porcine circovirus
ORGANISM Viruses: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hameir,A.L., Lin,L.L. and Nayar,G.P.
TITLE Nucleotide sequence of porcine circovirus associated with
postweaning multisystemic wasting syndrome in pigs
J. Virol. 72 (6), 5262-5267 (1998)
JOURNAL 98241772
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hameir,A.L., Lin,L.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services
Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,
Manitoba R3T 5S6, Canada
FEATURES
source Location/Qualifiers
1..1768
/organism="porcine circovirus"
/strain="pmws PCV"
/db_xref="taxon:46221"
/note="both strands of seven overlapping PCR fragments
were sequenced; virus isolated from lung, lymph node,
spleen and tonsil tissue from pigs affected by post
weaning multisystemic wasting syndrome"
51..995
/note="ORF1; similar to Rep protein encoded by
non-pathogenic PCV, GenBank Accession Number U49186;
predicted 35.8 kDa protein"
/codon_start=1
/product="putative Rep protein"
/protein_id="AAC59462.1"
/db_xref="PID:g2689646"
/db_xref="GI:2689646"
/translation="MPSKNGRSGPQHKRWFTLNPNSEDERKKIRELPISLFDYFI
VGEEGNEGRTPHLOGFANFKVKTQFNKVFAGCHIEKAKGTDOONKEYCSKEGN
LLIECGAPRSGQSDLSSTAVSLLESGSLVTVAEQHPVTFVRNFRGLAELKVSQGM
OKRDKNTNHHVIGVPPCGSKKWAANFADPTTYWKPKNKWDGYHGEVVIIDDFY
GWLPPDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSSSTAVPAVEALYR
RITSLVFNKNAEQSTEGGQFVTLSPCPPEFYEINI"
117..125
/note="glycosylation site"
327..332
complement(357..671)
/note="ORF3; predicted 11.9 kDa protein"
/codon_start=1
/protein_id="AAC59464.1"

```
/db_xref="PID:g2689648"
/db_xref="GI:2689648"
/translation="MTVPLPLVSWFPCGVCKKISSPFAFTTPRWPHNDVYIGLPI
TLHHPAHFQKFSQPAISDKRYRLCNGHQTALQOGTHSSROVTPLSRSSTF
NK"
CDS
complement(386..565)
/codon_start=1
/note="ORF4; predicted 6.5 kDa protein"
/protein_id="AAC59465.1"
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="MTCITLVQSRFCIFPLTFKSSAPRKFLTNVTCSSATVTRLPL
SNKVLTAVDRLSRCP"
misc_feature
complement(470..478)
/note="glycosylation site"
complement(688..753)
/note="ORF8; predicted 2.3 kDa protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHTVSDVDPHTAASHKSHQ"
misc_feature
816..824
/note="glycosylation site"
906..914
/note="glycosylation site"
983..988
complement(989..1033)
/note="ORF11; predicted 1.8 kDa protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKKNHYEVTKTQ"
polyA_signal
1016..1177
CDS
/note="ORF5; predicted 6.2 kDa protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="MVFIFHLGKVGKIFSELYIHGYTDIVVLVYVTFERSAEA
YVYHISRGL"
polyA_site
complement(1022..1027)
CDS
complement(1034..1735)
/note="ORF2; predicted 27.8 kDa protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="MTYPRRYRRRRHRRPSHLGQILRRRPWLHPHRYWRKNGI
FNRLSTFGYVKAATVPSPVAVDMFRNIDDFVPPGGTNKISIPFEYIRKVK
FEPWPCSPIQGGRGVSTAPVLDNDFVTATALYDYPVYNSRHTIPQPFYSHSRY
FTKPVLDSDIDYFOPNNKRTQLWLRLQTSRNVDRVGLGTAFENSIYQDYNIRVTY
VQREFNLKDPPLKP"
misc_feature
complement(1301..1309)
/note="glycosylation site"
complement(1522..1611)
/note="ORF6; predicted 3.1 kDa protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASPAPSDILSRLOSERPPGRWT"
CDS
1524..1631
/note="ORF10; predicted 4.1 kDa protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALTVYPKVRRVRVAKMPFFLLQR"
polyA_signal
1682..1741
CDS
/note="ORF7; predicted 1.9 kDa protein"
/codon_start=1
```

```
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGAVSSSAVTPPHIRHS"
complement(join(1732..1768,1..92))
/note="ORF9; predicted 4.6 kDa protein"
/codon_start=1
/protein_id="AAC59470.1"
/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSASSILLAGHVAAEVLPRCCRCRSALVILTAHFFRFQL"
stem_loop
join(1746..1768,1..13)
/function="putative replication site"
rep_origin
join(1762..1768,1..2)
/note="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
D49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN

alignment_scores:
  Quality: 1237.00      Length: 233
  Ratio: 5.402          Gaps: 0
  Percent Similarity: 98.283  Percent Identity: 96.996

alignment_block:
  US-09-209-961-5 x AF027217/rev
  Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgSe 17
1735 ATGACGCTATCAAGGAGCGGTACCGCAGAGAAGACACCGCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
1685 CCATCTGGCCAGATCTCCGCCGCCCTGCTGCTCCACCCCGCCGCC 1636
34 IsArgTyrArgTyrArgArgGlyAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGCTACCGCTGGAGAAGAAATGGCATCTTCAACACCGCCTCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
1585 CGCACCTTCGGATATACTGTCAAGGCTACCACAGTCAGACGCGCTCTG 1536
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
1535 GCGCGTGGACATGATGAGATTATATATTGAGACTTTGTCCTCCCGGAG 1486
84 lYgLYThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGACCAACAATCTCTATACCTTTGATACTACAGATAAGAAG 1436
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgGl 117
1435 GTTAAGGTTGAATTCGGCGCTGCTCCCATCACCCAGGCGTATAGGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
1385 AGTGGGCTCCAGCTGCTTTATCTAGATGATAACTTTGTAAACAAGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCCATATGTAACACTCTCTCCGCCATACATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
1285 CCCCAACCCCTCTCCTACCACTCCCGTTACTTCACACCAACCTGTCT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
```

```

1235 TGACTCCACTATTGATTACTTCCCAACCAATAACAAAAGGACTCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGAGGCTACAACCTCTAGAATGTGGACACGAGTGGCTCGGCACT 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 CGGTTCGAACACATATATACGACGACTACATATCCGTGTAAACCAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAGATTTAATCTTTAAGACCCCACTTAACCC 1037
|||||

seq_name: gb_v1:AF112862

seq_documentation_block:
LOCUS AF112862 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1768)
VIRUSES; ssDNA viruses; Circoviridae; Circovirus;
Hamel,A.L. and Nayyar,G.P.S.
Nucleotide sequence of four different isolates of porcine
circovirus detected in pigs with various clinical syndromes
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 1768)
Hamel,A.L. and Nayyar,G.P.S.
Direct Submission
TITLE
Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
JOURNAL
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/note="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
13..35
repeat_region
/ipt_type=tandem
/ipt_unit=13..18
CDS
51..995
/note="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:g4106918"
/db_xref="GI:4106918"
/translation="MPSRKNRSGPQPHKRWFTLNSSDEKRRKIRELPISLFDYFI
VGEENEGRTPLHGFANFKQTFNKVYLGARCHEKAGTDQONKEYGSKGN
LIECGAPRSQGRSDLSAVTSLGSLVTVAEQHPVTFVRNFRGLAELLVSGKM
QKRDWNTYHVIVGPPCGKSKWAANFADPTTYKPPRNKWDYGHGEVNVDDFY
GWLPHDDLRLCDRIPLTVETKGTVPFLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFWRNATEQSTEEGGQFVTLSPPCPEFFPEYNI"
327..332
polyA_signal
complement(357..671)
CDS
/codon_start=1
/product="ORF-3"
/protein_id="AAD03087.1"
/db_xref="PID:g4106919"
/db_xref="GI:4106919"
/translation="WYTIPLVSRWFPVCGRVCKYKISSPFAFTTPRPHNDVDIGLPI
TLLHFFAHQKFSQPAEISDKRYRVLLCNGHQTALQQGTHSSRQVTPLSLRSSTF
NK"
complement(386..565)
CDS
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD03085.1"
/db_xref="PID:g4106917"
/db_xref="GI:4106917"
/translation="MTPYRRYRRRRRSHLQILRRRPWLVPYRRYVWRKNGKI
FNRLSRTGYTVKATVTTPSWAVDMRENIDDFVPPGGTNRKISYVPEYIRKVK
VEFWPCSPITGGDRGVGSTAVILDDNFVTATALTIDFYVNISSRHIPQFFSHSR"
complement(1522..1611)
CDS
/codon_start=1
/product="ORF-6"
/protein_id="AAD03084.1"
/db_xref="PID:g4106916"
/db_xref="GI:4106916"
/translation="MASSTPASPAPSDILSRLPOSQRPGRWT"
1524..1631
CDS
/codon_start=1
/product="ORF-10"
/protein_id="AAD03092.1"
/db_xref="PID:g4106924"
/db_xref="GI:4106924"
/translation="MSTAQEGVTVWALTVPYKVRERRVLKMPFFLLQR"
1682..1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD03093.1"
/db_xref="PID:g4106925"
/db_xref="GI:4106925"
/translation="MAAGAVSSSAVTPPWIRHS"
join(1750..1768,1..13)
note="putative replication site"
stem_loop

```

rep_origin join(1762...1768,1..2)
/note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186"

BASE COUNT 448 a 360 c 493 g 467 t
ORIGIN

alignment_scores:

Quality: 1237.00 Length: 233
Ratio: 5.402 Gaps: 0
Percent Similarity: 98.283 Percent Identity: 96.567

alignment_block:

US-09-209-961-5 x AF112862/rev ..

Align seg 1/1 to reverse of: AF112862 from: 1 to: 1768

```
1 MetThrTyrProArgArgTyrArgArgArgArgHisArgProArgSe 17
1735 ATGCGTATCCAGGAGCGGTACCGCAGAGAGACACCGCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
1685 CCATCTTGGCCAGATCTCCGCCCGCGCCCTGGCTGCTCCACCCGCC 1636
34 isArgTyrArgTTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGCTACCGTGGAGAGGAAATGCAATCTTCAACACCGCCCTCC 1586
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
1585 CGCACCTTCGGATATCTGTCAAGGCTACACACGTCACACGCCCTCTG 1536
67 pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
1535 GCGCGTGGACATGATGATTAATATTGACGACTTGTCCCGCGGAG 1486
84 lGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
1485 GGGGGACCAACAAATCTGTACCTTTGAATACTACAGAATAAGAAG 1436
101 ValLysValGluPheTppProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAGTTGAATCTCGCCCTGCTCCCCATCACCGAGGTGATAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTGGGCTCCACTGCTGTTATCTAGATGATAACTTTGTAACAAAGCCA 1336
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACTATGACCCATATATTAATACTACTCTCCCGCATACAATC 1286
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValLe 167
1285 CCCAACCCCTTCCTACCACTCCCGTTAATTCACCCAAACCTGTACT 1236
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGACAGAACTATTGATTACTTCCAAACCAATAACAAAAAATCAGCTTT 1186
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GCGTGGCTACAAACCTCTAGAAATGTGGACACGCTAGGCTCGGCAC 1136
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTCGAAACAGTAATAATACGACGAGACTACATAATATCCGTGTAA 1086
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACAATTCAGAAATTAATCTTAAAGACCCCTTAAACCC 1037
```

seq_name: gb_vi:AF147751

seq_documentation_block:

LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751

NID 95007010

VERSION AF147751.1 GI:5007010

KEYWORDS porcine circovirus type 2.

SOURCE porcine circovirus type 2

ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and Sorden, S.

TITLE Characterization of clinical and immune responses in young swine to experimental porcine circovirus type II infection.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 1768)

AUTHORS Yoon, K.-J. and Pogranichnyy, R.M.

TITLE Direct Submission

JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa

State University, College of Veterinary Medicine, Ames, Iowa 50011,

USA

FEATURES Location/Qualifiers

source

1..1768 /organism="porcine circovirus type 2"

/isolate="ISUVDL 98-15237"

/specific_host="Sus scrofa"

/db_xref="taxon:85708"

/note="isolated from pigs with postweaning multisystemic wasting syndrome"

51..995

/note="ORF1: 35-36kd"

/codon_start=1

/product="unknown"

/protein_id="AAD37776.1"

/db_xref="PID:g5007011"

/db_xref="GI:5007011"

/translation="MPKSKNGRSGPQPKRWVFTLNNPSEDERKKIRELPISLDFYFI

VGEENEGRTPLHGFANFVKQTFNKVWYLGARHIEKAKGTDOONKEVCKSEGN

LLIECGAPRSGORSDLSTAVSLLESGSLVTVAEQHPVTFVNRGLAETLLKYSGM

OKRDWKNVHVIVGPPCCGSKNAANFADPTTYWKPRKNWDCYHGEVWVDDFY

GWLPWDDLRLCDRIPLTVETKGTVPFLAKSLITISNQTPLEWISSTAVPAVEALYR

RITSLVFWKNATQSEEGGQVFTLSPPCPEFFEYIN"

complement(1034..1735)

/note="ORF2: 27-28kd"

/codon_start=1

/product="unknown"

/protein_id="AAD37777.1"

/db_xref="PID:g5007012"

/db_xref="GI:5007012"

/translation="MTYPRRRYRRRRPRSHLGLQLRRRPWLVPRRYVRRKNGI

ENRLSTFGYTVKATVTPSPWAVMMRENIDFVPPGGTNRKISIPFEYRKRKVK

VEFWPCSPITGGDRGVSTAVILDNFVTATALTIDPYVNSRRHTIPQPFSCHSRY

FTPKPVLDSIDTYFQPNKRNQLWLRLQTSRNVHDVGLGTAFENSIDYQDYNIRVTY

VQFREFNLKDPPLKP"

BASE COUNT 450 a 362 c 494 g 462 t

ORIGIN

alignment_scores:

Quality: 1234.00 Length: 233
Ratio: 5.412 Gaps: 0
Percent Similarity: 97.854 Percent Identity: 96.996

alignment_block:

US-09-209-961-5 x AF147751/rev ..

Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768

1 MetThrTyrProArgArgTyrArgArgArgHisArgProArgSe 17

```
|||||
1735 ATGACGTATCAAGGAGCGTTACCGCAGAGAGACACCGCCCGCAG 1686
17 rHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTCGGCAGATCTCCGCGCGCCCTGGCTGCTCCACCGCCGCC 1636
34 isArgTyArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGCTACCGCTTGAGAGGAAATGATGATCTTCAACACCGCGCTCC 1586
51 ArgThrPheGlyTyThrValLysArgThrThrValThrTrpSerTr 67
|||||
1585 CGCACCTTCGGATATCTGTCAGGCTACACAGTCAGACGCGCTCATG 1536
67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
|||||
1535 GCGCGTGCATGATGATGATTAATATGACGACTTGTTCGCCCGGAG 1486
84 LyGlyThrAsnLysIleSerIleProPheGlyTyThrArgIleArgLys 100
|||||
1485 GGGGAGCAACAAATATCCATCCCTTGAATATCTACAGAAATAGAAAG 1436
101 ValLysValGluPheTTPProCysSerProIleThrGlnGlyAspArgG1 117
|||||
1435 GTTAAGTTGNAATCTGGCCCTGCTCCCGCATCACCGAGGTATAGGG 1386
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGGCTCCACTGCTGTATTCTAGATGATAAATTTGTAACAAAGCCA 1336
134 hrAlaLeuThrTyAspProTyfValAsnTyfSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACTATGACCCATATGTAATCTACTCTCCCGCATCAATC 1286
151 ProGlnProPheSerTyfHisSerArgTyfPheThrProLysProValIe 167
|||||
1285 CCGCAACCTTCTCTGCCACTCCGTTACTTTCACACCCAAACCTGTCT 1236
167 uAspSerThrIleAspTyfPheGlnProAsnAsnLysArgAsnGlnLeu 184
|||||
1235 TGACTCCACTATTGATTACTTCCAAACCAATAACAAAGGAATCAGCTT 1186
184 rPleuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGGAGGCTACAACCTCTAGAATGTGGACACGAGTGGCGCAG 1136
201 AlaPheGluAsnSerLysTyfAspGlnAspTyfAsnIleArgValThrMe 217
|||||
1135 CGGTTCGAAACAGATATATACGACGAGCTACATATATCGGTGAACCAT 1086
217 tyfValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAGAATTAACTTTAAAGACCCCGCCCTTAAACC 1037
seq_name: gb_v1:AF109397
```

```
seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE
ORGANISM
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
JOURNAL
Unpublished
REFERENCE .2 (bases 1 to 1768)
```

```
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE
3 (bases 1 to 1768)
AUTHORS
Hamel,A.L. and Nayar,G.P.S.
TITLE
Direct Submission
JOURNAL
Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REMARK
Sequence update by submitter
COMMENT
On Feb 2, 1999 this sequence version replaced gi.4028609.
FEATURES
Location/Qualifiers
1..1768
/organism="bovine circovirus"
/virion
/db_xref="taxon:85542"
/notes="BCV
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217."
13..35
/repeat_region
/rpt_type=tandem
/rpt_unit=13..18
51..995
/notes="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AADI1928.1"
/db_xref="PID:g4210314"
/db_xref="GI:4210314"
/translation="MPSKNGRSGPOPHKRWFTLNNPSDEKRIKRELPSLFDYFI
VGEENEGTPTHLGFANFVKQTKNKVYLGARCHIKAKTDOQKEYCSKEGN
LLIECGAPRSGQSRSLSTAVSLLESGLTVVAEQHPYTFVFRGLAEALLKVSQKM
QKRDWNTNHHVIVGPPCGSKWAANFADPTTYWPKRNKWDGYHGEVYVDDFY
GWLWDLLRLCDRPLTVETGKTPVFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEFFPEYNY"
327..332
polyA_signal
complement(357..671)
/notes="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AADI1930.1"
/db_xref="PID:g4210315"
/db_xref="GI:4210315"
/translation="MVTIPPLVSRNFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
TLILFFPAHFQKFSQPAEISDKRYRVLNCGHQTALQOQGHSSRQVTLRLSRSTF
NK"
complement(386..565)
/notes="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AADI1931.1"
/db_xref="PID:g4210316"
/db_xref="GI:4210316"
/translation="MCTLVFQSRFCIFPLTFKSSASPRKFLTNVTCSSATVTRLPL
SNKVLTAVDRLSLRCP"
553..732
CDS
/codon_start=1
/product="ORF-12"
/protein_id="AADI1927.1"
/db_xref="PID:g4208610"
/db_xref="GI:4028610"
/translation="MYTSLMGHLGVKANGLLILQTRKPTGHNHLETSGGMYTMVKW
LLMPTFMAGCRGMIV"
complement(688..753)
CDS
/codon_start=1
```

```

/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:g4028615"
/db_xref="GI:4028615"
/translation="WVEIHLGFKWGVKIFSELYIHGYTDIVVLVVTTFERSAEA
YVNHISRL"
complement(1022..1027)
complement(1236..1735)
/notes="Similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:g4028612"
/db_xref="GI:4028612"
/translation="MTVPRRRYRRRRTPRSHLGILRRRPWLHPHRYRWRKNGI
FNRLSRFTGVKATVTPSWADMMREINDDFVPGGTNKISIPFXYRIRKVK
VEFWPCSPITGDRGVSTAVILDDNFVTKALTYDPVYNSRHTIPQPSYHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASSTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="MSTAQEGVLTVALTYPKVRERVLKMPFELLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:g4028617"
/db_xref="GI:4028617"
/translation="MAAGAGPSSAVTPPWIRHS"
join(1750..1768.1..13)
join(1762..1768.1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U9186"
452 a 362 c 493 g 461 t

BASE COUNT
ORIGIN

alignment_scores:
  Quality: 1230.00      Length: 233
  Ratio: 5.395          Gaps: 0
Percent Similarity: 97.854 Percent Identity: 96.996

alignment_block:
US-09-209-961-5 x AF109397/rev ..

Align seg 1/1 to reverse of: AF109397 from: 1 to: 1768

```

```

1 MetThrTyrProArgArgTyrArgArgArgHisArgProArgSe 17
|||||
1735 ATGACGATCCAGGAGCGCTTACCGCAGAGAGAGAGAGAGAGAG 1686
|||||

17 rHisLeuGlyGlnIleLeuArgArgProTyrLeuValHisProArgH 34
|||||
1685 CCATCTTGGCAGATCTCCGCCGCCCTGGCTGCTGCACCCGCC 1636
|||||

34 isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGCTACCGTGGAGAGAGAAATGGCATCTTCAACACCCGCCCTCTCC 1586
|||||

51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTi 67
|||||
1585 CGACCTTCGGATATCTGTCAAGGCTACACAGTCAGACGCCCTCTCTG 1536
|||||

67 palavalAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1485 GGGGGACCAACAAATCTATACCTTTGAATACTACAGATAAGAAAG 1436
|||||

101 ValLysValGluPheTyrProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGGTGAATCTGGCCCTGCTCCCATCATCCAGGGGTATAGGGG 1386
|||||

117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
|||||
1385 AGTGGGCTCCACTGCTGTTATCTAGATGATAACTTTGTAAACAAGGCCA 1336
|||||

134 hralalaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCCATATGTAACACTCTCTCCGCCCATACAATC 1286
|||||

151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
|||||
1285 CCCAACCCCTCTCTACCATCTCCGTTAATTCACCCCAACCTGTCT 1236
|||||

167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGACTCCACTATTGATTACTTCCACCAATATACAAAGGAATCAGCTTT 1186
|||||

184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGAGGCTACAAACCTCTAGAAATGTGGACCACGTAGGCTCGGCAC 1136
|||||

201 AlapheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCCTTGGAAACAGTAAATACGACGAGCTACAAATATCCGTGTAAACAT 1086
|||||

217 tTyrValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAAATTAATCTTAAGAGACCCCCCACTTAACCC 1037
|||||

```

seq_name: gb_vi:AF154679

seq_documentation_block:

LOCUS AF154679 1768 bp DNA circular VRL 13-JUN-1999
 DEFINITION Porcine circovirus, complete genome.

ACCESSION AF154679

NID g5052004

VERSION AF154679.1 GI:5052004

KEYWORDS

SOURCE porcine circovirus.

ORGANISM

REFERENCE

1 (bases 1 to 1768)

AUTHORS Kuo,T.Y., Chlou,Y.C. and Lai,S.S.

TITLE Complete nucleotide sequences analysis of porcine circovirus

JOURNAL outbreak in Taiwan

UNPUBLISHED

```
REFERENCE 2 (bases 1 to 1768)
AUTHORS Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE Direct Submission
JOURNAL Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan
University, 142, Chousan Road, Taipei 106, Taiwan
FEATURES
source
1..1768
/organism="porcine circovirus"
/isolate="MLW98"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/notes="putative Rep protein"
/codon_start=1
/product="ORF1"
/protein_id="AAD38398.1"
/db_xref="PID:g5052005"
/db_xref="GI:5052005"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKRELPLISLDFCFI
VGEENEGRTPLHQFANFKVKTENKVKWYGARCHIEKAKGTDQONKEYCSKEGN
LLIECCAPRSQGRSLSSTAVSLTSEGSLVTVAEQHPVTFVRNFRGLAELLKVSQGM
QKRDWNTNHHVIVGPPCGSKWAANFADPTTYWPKRNKWDGYHGEVWIDDFY
GWLPMDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATQSTEEGGQFVTLSPPCPEFFEYIN"
BASE COUNT 451 a 367 c 495 g 455 t
ORIGIN

alignment_scores:
Quality: 1226.00 Length: 233
Ratio: 5.330 Gaps: 0
Percent Similarity: 98.712 Percent Identity: 95.708

alignment_block:
US-09-209-961-5 x AF154679/rev ..
Align seg 1/1 to reverse of: AF154679 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgGHisArgProArgSe 17
|||||
1735 ATGCGTATCAAGGAGCGGTTTCGCGAGCAGGAGACACCGCCCGCGAC 1686
|||||
17 rHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgH 34
|||||
1685 CCATCTTGCCAGATCCTCCGCGCGCGCCCTGCGTCTGCCACCCCGCC 1636
|||||
34 isArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 AGCGTTACCGCTGGAGAGGAAATGCGATCTTCACACCGCGCTCTCC 1586
|||||
51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
|||||
1585 CGCACCTTCGGATATACGTCAAGGCTAGCAGCAGTACAGAACGCCCTCTG 1536
|||||
67 pAlaValAspMetArgPheLysIleAspAspPheValProGlyG 84
|||||
1535 GCGCGTGACATGAGATTTGATATTAAACGACTTGTTCCTCCCGGGAG 1486
|||||
84 lYgIlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCAACAAATCTCTATACCTTTGATACACTACAGATNAGAAG 1436
|||||
101 ValLysValGluPheTrpProCysSerProIleThrGlnGlyAspArgG 117
|||||
1435 GTTAAGGTGAATTCGCGCCTGCTCCCAATCACCACCGGTGACAGGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
|||||
1385 AGTGGGCTCCACTGCTGTATTCTAGATGATAACTTGTGAACAGGCCA 1336
|||||
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerArgHisThrIle 150
|||||
1335 CAGCGCTAACCTATGACCCCTATGTAACACTCTCTCCCGCCCATCATC 1286
```

```
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
|||||
1285 CCCACACCTCTCTTACCACCTCCCGGTACTTACCCCAACCTGTCTCT 1236
|||||
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
|||||
1235 TGATCCACTATTGATTACTTCCACCAACAACAAGAAAGATCAGCTTT 1186
|||||
184 rPleuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GCGTGGGCTACAAACCTCGCAATGTGGACACGCTAGGCTCGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 GCGTTCGAAACAGTAATAACACAGGACTACATATATCCGTGTAAC 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1085 STATGTACATTCAGAGATTTAATCTTAAAGACCCCACTTAAACCC 1037
|||||
seq_name: gb_vi:AF109399
seq_documentation_block:
LOCUS AF109399 1768 bp DNA circular VRL
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID 94106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE
ORGANISM porcine circovirus type 2-E.
Viruses; ssDNA viruses; Circoviridae; Circovirus; #69#ine
circovirus type 2.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1..1768
/organism="porcine circovirus type 2-E"
/virus
/db_xref="taxon:85544"
/notes="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
/notes="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKNGRSGPQPHKRWFTLNNPSEDERKKRELPLISLDFYI
VGEENEGRTPLHQFANFKVKTENKVKWYGARCHIEKAKGTDQONKEYCSKEGN
LLIECCAPRSQGRSLSSTAVSLTSEGSLVTVAEQHPVTFVRNFRGLAELLKVSQGM
QKRDWNTNHHVIVGPPCGSKWAANFADPTTYWPKRNKWDGYHGEVWIDDFY
GWLPMDDLRLCDRYPLTVETKGTVPFLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFNKATQSTEEGGQFVTLSPPCPEFFEYIN"
327..332
poly_a_signal
```



```
184 rpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGATGAGACTCCAAACAGTAGAATGTGGACCACTAGGCTCGGCAC 1136
|||||
201 AlaPheGluAsnSerLysTytrAspGlnAspTytrAsnIleArgValThrMe 217
|||||
1135 CGGTGCGAAACAGTAATACGACGAGCTACATAATCCGTGTAACCAT 1086
|||||
217 tytrValGlnPheArgGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1085 GTATGTACAATTACAGAGATTAACTTAAGACCCCCCCTAAACCC 1037

seq_name: gb_vi:AF109398

seq_documentation_block:
LOCUS AF109398 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-C, complete genome.
ACCESSION AF109398
NID 94106885
VERSION AF109398.1 GI:4106885
KEYWORDS
SOURCE
ORGANISM
VIRUSES: ssDNA viruses; Circoviridae; Circovirus; #2##line
circular type 2.
REFERENCE
1 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 1768)
Hamel, A.L. and Nayar, G.P.S.
Direct Submission
TITLE
Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1. 1768
/organism="porcine circovirus type 2-C"
/virus
/db_xref="taxon:85543"
/note="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to Porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-C
designation is based upon restriction endonuclease
digestion pattern"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
CDS
51..995
/note="ORF-1."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03061.1"
/db_xref="PID:94106886"
/db_xref="GI:4106886"
/translation="MPSKNGRSGQPQPKRWVFTLNNSSEDERKKIREPLISFDYFI
VGEENEGRTPHLQGFANFKVKTQFNKVYLGARCHIEKAKGTDQONKEVCSKEG
LLQCCGPRSGQSRDLSAVSLLESGLTVVAEQHPVTFVFRNGLAELLKVSQGM
QKRDWNTNHIIVGPGCGSKAANFADPETTYTKPKRNKWDGTYGHEEVVYIDDFY
GWLPPDLDLCLCDPLTYVETKGGPVFLARSILITNSQPLEYSSYTAVPAAVLYR
RITSLVFNKNAEQSTEEGGQVFTLSPCCPEFFEYNI"
polyA_signal
327..332
complement(357..671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03063.1"
/db_xref="PID:94106888"
/db_xref="GI:4106888"
/translation="MTYIPPLVSRWFVPCGRVCKISSPFAFATPRPHNDVYIRLPI
TLHFPAHFQKFSQPAISDKRYRLVLCNGHQTPALQQGHSSQVTALESRSSTL

184 complement(386..565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03064.1"
/db_xref="PID:94106889"
/db_xref="GI:4106889"
/translation="MTCTFFVQSRFCIFPLTFKSSASPRKFLTNVTCGCSATVTRLPL
SNKVLTAVDRLRCP"
CDS
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03070.1"
/db_xref="PID:94106895"
/db_xref="GI:4106895"
/translation="MYTSLWGHGLGVAKANGLLILQTRKPHGTGNHLETSGGMVTVVKWK
LLLMFMVGCGRMIY"
polyA_signal
983..988
complement(989..1033)
CDS
/codon_start=1
/product="ORF-11"
/protein_id="AAD03069.1"
/db_xref="PID:94106894"
/db_xref="GI:4106894"
/translation="MNKNHVEVIKKTQ"
CDS
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD03065.1"
/db_xref="PID:94106890"
/db_xref="GI:4106890"
/translation="MVPIHLGFKMGVFKIKFSELYIHGYTDIVVLVYTVFERSAEA
IMVYISSSL"
polyA_signal
complement(1022..1027)
CDS
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD03062.1"
/db_xref="PID:94106887"
/db_xref="GI:4106887"
/translation="MTYPRRYRRRTTSPRSHLGOILRRRPLVHPHRYBWRKNGI
FNTRLSFTFGTVATVTRTSWADMLRFKIDDFVPCGGTNKISIPFEYIRKVK
VEFWPCSPITQDGRGVGSSAVILDDNFVKATALTYDPYVNYSSHTITQPFVSHSR"
CDS
complement(1519..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03066.1"
/db_xref="PID:94106891"
/db_xref="GI:4106891"
/translation="MASSTPASPAPSDILSLPQSERPPGRWTC"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03068.1"
/db_xref="PID:94106893"
/db_xref="GI:4106893"
/translation="MSTAQEGVLTVVLTVPYKVRERRVLKMPFFELLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03067.1"
/db_xref="PID:94106892"
/db_xref="GI:4106892"
/translation="MAAGAPSSAVTPPWIRHS"
stem_loop
join(1750..1768,1..13)
/note="putative replication site"
rep_origin
join(1762..1768,1..2)
/note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186"
BASE COUNT 449 a 364 c 490 g 465 t
```

```
ORIGIN
alignment_scores:      Quality: 1207.00      Length: 233
                       Ratio: 5.341        Gaps: 0
Percent Similarity: 96.996 Percent Identity: 94.850

alignment_block:
US-09-209-961-5 x AF109398/rev ..

Align seg 1/1 to reverse of: AF109398 from: 1 to: 1768

1  MetThrTyrProArgArgArgTyrArgArgArgGHisArgProArgSe 17
|||||
1735 ATGACGTATCAAGGAGGCGTTACCGCAGAAGAAGACCGCGCCGCGAG 1686
|||||
17  rHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProArgH 34
|||||
1685 CCATCTTGCCAGATCCTCCGCGCGCGCCCTGGCTGCTCCACCCCGCC 1636
|||||
34  isArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
|||||
1635 ACCGTTACGCTGGAGAAAGAAATGGCATCTTCAACACCGCCCTCTCC 1586
|||||
51  ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
|||||
1585 GCGACCTTCGGATATACGTCAAGGCTACACAGTCAGAACGCCCTCCGTG 1536
|||||
67  pAlaValAspMetMetArgPheLysIleAspPheValProProGlyG 84
|||||
1535 GCGGTGACATGCTGAGATTAAATTAACGACCTTGTCCCGCAGGAG 1486
|||||
84  lYgLYThrAsnLysIleSerIleProPheGluTyrTyrArgIleArgLys 100
|||||
1485 GGGGACCAACAATACTCTATACCTTTGAATACTACAGAATAAGAAAG 1436
|||||
101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgGl 117
|||||
1435 GTTAAGGTGAATCTGCGCCCTGCTCCCAATCACCGGGTGACAGGGG 1386
|||||
117 yValGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAlaT 134
|||||
1385 AGTGGATCCAGTGTCTATCTAGATGATAACTTTGTACCAAGGCCA 1336
|||||
134 hrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
|||||
1335 CAGCCCTAACCTATGACCCCTATGTAACCTACTCTCCCGCCATACCAT 1286
|||||
151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
|||||
1285 ACCGAGCCCTCTCTACACACTCCCGGTTAATTAACACCCAAACCTGCT 1236
|||||
167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeuT 184
|||||
1235 TGATGGGCAATCGATTACTTCCAAACCAATTAACAAAGAAATAACTCT 1186
|||||
184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
|||||
1185 GGCTGAGACTACAACTACGAAATGTAGACCATGTAGCCCTCGGCACAT 1136
|||||
201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
|||||
1135 CGGTTCGAAAACAGTATATACCGCAGGACTACAAATATCCGTGTAAACCAT 1086
|||||
217 tTyrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1085 GTATGTACAATTCAGAGAATTTAATCTTAAAGACCCCGCCACTTAAACCC 1037
|||||

seq_name: gb_vi:AF117753
seq_documentation_block:
LOCUS      AF117753      1768 bp      DNA      circular      VRL      04-FEB-1999
```

```

/product="ORF-8"
/protein_id="AAD12312.1"
/db_xref="PID:g4219098"
/db_xref="GI:4219098"
/translation="MDTDHTVSVVDHTAASHKSHQ"
983..988
polyA_signal
CDS
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD12313.1"
/db_xref="PID:g4219099"
/db_xref="GI:4219099"
/translation="MNNKNHYEIKKTQ"
1016..1177
CDS
/codon_start=1
/product="ORF-5"
/protein_id="AAD12314.1"
/db_xref="PID:g4219100"
/db_xref="GI:4219100"
/translation="MVFIHLGFKWGVFKIKFSELYIHGYTDIVLVVVYTVFERSAEA
YMWYISSSL"
1022..1027)
polyA_signal
CDS
complement(1026..1735)
/notes="similar to Porcine circovirus ORF-2 product encoded
by the sequence presented in GenBank Accession Number
AF027217"
/codon_start=1
/product="ORF-2"
/protein_id="AAD12315.1"
/db_xref="PID:g4219101"
/db_xref="GI:4219101"
/translation="MTFFRRYRRRRRSHLGOILRRPFWELHRRHVRWRRKNCI
FNTLRSGTYVVKATVTPSWADMLRKFIDDFVPGGKNKISIPFYIRKVK
VEFWPCSPITQGDGRGVSSAVILDDNFFPKFTALTYDPYVYSSRHTITOPFSYHSR"
1524..1631
CDS
/codon_start=1
/product="ORF-10"
/protein_id="AAD12317.1"
/db_xref="PID:g4219103"
/db_xref="GI:4219103"
/translation="MSTAQEGVLTVVVALTYPKVRRRLVKMPFFLLQR"
complement(1528..1611)
CDS
/codon_start=1
/product="ORF-6"
/protein_id="AAD12316.1"
/db_xref="PID:g4219102"
/db_xref="GI:4219102"
/translation="MASSTPASPAPSDILSRLPQSRPPGR"
1682..1741
CDS
/codon_start=1
/product="ORF-7"
/protein_id="AAD12318.1"
/db_xref="PID:g4219104"
/db_xref="GI:4219104"
/translation="MAAGAVSSSAVTPPKRHS"
join(1750..1768,1..13)
stem_loop
/notes="putative replication site"
join(1762..1768,1..2)
rep_origin
/notes="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF027217
and U49186."
452 a 358 c 497 g 461 t

BASE COUNT 452 a 358 c 497 g 461 t
ORIGIN

alignment_scores:
  Quality: 1202.00      Length: 233
  Ratio: 5.319          Gaps: 0
  Percent Similarity: 96.996      Percent Identity: 93.562

alignment_block:
  US-09-209-961-5 x AF117753/rev ..

```

```

Align seg 1/1 to reverse of: AF117753 from: 1 to: 1768

1 MetThrTyrProArgArgArgTyrArgArgArgArgHisArgProArgSe 17
1735 ATGACGTTTCCAAGGAGGGGTTACCGCAGACGACGACCCCGCCGAG 1686

17 rHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProArgH 34
1685 CCATCTTGCCAGATCCCTCCGCGCGCCCTGGTTCTCTCCACCCCGCC 1636

34 IsArgTyrArgTyrArgArgLysAsnGlyIlePheAsnThrArgLeuSer 50
1635 ACCGTTACCGCTGGAGAGGAAAAATGGCATCTCAACACCCGCTCTCC 1586

51 ArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSerTr 67
1585 CCGACCTTCGGATATCTCTCAAGGCTACACAGTCAGAACGCCCTCTG 1536

67 pAlaValAspMetMetArgPheLysIleAspAspPheValProGlyG 84
1535 GCGGTAGACATGCTGAGATTAAATATAGACGACTTTGTTCCCGGGAG 1486

84 IyGlyThrAsnLysIleSerIleProPheGlyTyrTyrArgIleArgLys 100
1485 GGGGACCAACAAATCTCTATACCTTTGAATACTACAGATAAGAAAG 1436

101 ValLysValGluPheThrProCysSerProIleThrGlnGlyAspArgG 117
1435 GTTAAGGTTGAATCTGCGCCTCTCCCAATCACCCAGGTGACAGGGG 1386

117 vAlaGlySerThrAlaValIleLeuAspAspAsnPheValThrLysAla 134
1385 AGTTGGATCCAGTGTCTATTCTAGATGATACTTTTCCCAAGAGCA 1336

134 hAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHisThrIle 150
1335 CAGCCCTAACCTATGACCCCTATGAAACTACTCTCCCGCCATACCATA 1286

151 ProGlnProPheSerTyrHisSerArgTyrPheThrProLysProValle 167
1285 ACCCAACCCCTTTCTACCACTCCGCTTAATTACACACCCCAACCTGCT 1236

167 uAspSerThrIleAspTyrPheGlnProAsnAsnLysArgAsnGlnLeu 184
1235 TGATTCGACATCGATTACTTCCAGCCCAATAACAAAGAAATCAACTCT 1186

184 rPLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuGlyThr 200
1185 GCCTGAGACTACAACTACTGGAATGTAGACCATGTAGCCCTCGCACT 1136

201 AlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgValThrMe 217
1135 GCGTTCGAAACAGTATATACGACGAGGACTACATATATCCGTGTAAACC 1086

217 tInrValGlnPheArgGluPheAsnLeuLysAspProProLeuGluPro 233
1085 GTATGTACAATTACAGAAATTAATCTTAAAGACCCCCCACTTAAACCC 1037

seq_name: gb_v1:AF118097

seq_documentation_block:
  LOCUS AF118097 700 bp DNA VRL 27-JUL-1999
  DEFINITION Porcine circovirus type 2 isolate IAF-4370 unknown gene.
  ACCESSION AF118097
  NID 95596427
  VERSION AF118097.1 GI:5596427
  KEYWORDS porcine circovirus type 2.
  SOURCE porcine circovirus type 2
  ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
  REFERENCE 1 (bases 1 to 700)
  AUTHORS Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.
  TITLE Multiplex PCR for detection and typing of porcine circoviruses

```

J. Clin. Microbiol. (1999) In press
REFERENCE 2 (bases 1 to 700)
AUTHORS Ouardani, M., Wilson, L., Jette, R., Montpetit, C. and Dea, S.
TITLE Direct Submission
JOURNAL Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des
Prairies, Laval, Que H7N 4Z3, Canada
FEATURES
source Location/Qualifiers
1..700
/organism="porcine circovirus type 2"
/isolate="IAF-4370"
/db_xref="taxon:85708"
/note="virus isolated from lung, lymph node, spleen and
tonsil tissue from pigs affected by postweaning
multisystemic wasting syndrome"
CDS complement(42..>700)
/note="ORF2"
/codon_start=3
/product="unknown"
/protein_id="AAD45581.1"
/db_xref="PID:g5596428"
/db_xref="GI:5596428"
/translation="RSLGQILRRPWLHPHRYRWRKNGIENRLSRFTGYTKA
TVRTPSWADMRNFDDFVPPGGGKNSIPFEYRIRKVKVEFWPCSPITQDRG
VGSTAVLDNFTKATLYDPVYNNSSRHTIPQFYSHSRYFTPKPVLDSITDYFQ
PNNKRNQLMRLQTSRVNDHVLGTAFAFENSKYDQDYNIRVTMYVQFREFNLKDPPLKP
"

BASE COUNT 156 a 127 c 210 g 207 t
ORIGIN

alignment_scores:
Quality: 1163.00 Length: 218
Ratio: 5.409 Gaps: 0
Percent Similarity: 98.624 Percent Identity: 97.248

alignment_block:
US-09-209-961-5 x AFL18097/rev ..

Align seg 1/1 to reverse of: AFL18097 from: 1 to: 700

16 ArgSerHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisPr 32
|||||
698 CGCAGCCATCTGGCAGATCTCCGCCGCCCGCTGGCTCGTCCACCC 649
32 oArgHisArgTyrArgTyrArgGlyAsnGlyIlePheAsnThrArgL 49
|||||
648 CCGACACCGCTACCGTGGAGAAGGAAAAATGGCATCTTCAACACCCGCC 599
49 euSerArgThrPheGlyTyrThrValLysArgThrThrValThrPro 65
|||||
598 TCTCCCGCACCTTCGGATATAGTCAAGGCTACACAGTCAAGACGCC 549
66 SerTrpAlaValAspMetMetArgPheLysIleAspAspPheValPro 82
|||||
548 TCCTGGCGGTGGACATGATGAGATTATATGACGACTTGTTCGCC 499
82 oGlyGlyThrAsnLysIleSerIleProPheGluTyrTyrArgIleA 99
|||||
498 GGGAGGGGGGACCAAAATCTCTATACCTTTGAATACTACAGAATAA 449
99 IgLysValLysValGluPheTrpProCysSerProIleThrGlnGlyAsp 115
|||||
448 GAAAGGTTAAGGTTCAATTCCTGGCCCTGCTCCGCCATCACCAGGTGAC 399
116 ArgGlyValGlySerThrAlaValIleLeuAspAspAsnPheValThrLy 132
|||||
398 AGGGAGTGGGCTCCACGCTGCTGTTATCTAGATGATACTTTGTAACAA 349
132 sAlaThrAlaLeuThrTyrAspProTyrValAsnTyrSerSerArgHist 149
|||||
348 GGCCACAGCCCTAACCTATGACCCCATATGATAACTACTCTCCCGCCATA 299
149 hrIleProGlnProPheSerTyrHisSerArgTyrPheThrProLysPro 165

|||||
298 CAATCCCCCAACCTTCTCTACCACTCCCGTTACTTTCACACCCAAACCT 249
166 ValLeuAspSerThrIleAspTyrPheGlnProAsnAsnLysAlaGsnG1 182
|||||
248 GTCTTGACTCCACTATGATTACTTCCAAACCAATACAAAAGGAATCA 199
182 nLeuTrpLeuArgLeuGlnThrSerGlyAsnValAspHisValGlyLeuG 199
|||||
198 GCITTTGATGAGGCTACAAACCTCTAGAAATGTGGACCACGTAGGCCTCG 149
199 lyThrAlaPheGluAsnSerLysTyrAspGlnAspTyrAsnIleArgVal 215
|||||
148 GCACCTGCTTCGAAACAGTAAATACGACGAGGACTACAAATATCCGTGTA 99
216 ThrMetTyrValGlnPheArgGluPheAsnLeuLysAspProLeuG1 232
|||||
98 ACCATGTATGTACAATTCAGAGAATTTAATCTTAAAGACCCCCACCTAA 49
232 upro 233
48 ACC 45
:|||


```
1517 CCTCTGGAATGTGACACGAGTAGCGCTCGGCTCGTGAACACAGT 1468
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1467 AAATACGACACGAGGTACAAATATCGGTGAACCATGTATGTACAATTCAG 1418
|||||
222 gGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1417 AGAATTTAATCTTAAAGACCCCTTAAACCC 1384
```

seq_name: N_Geneseq_36:X35012

seq_documentation_block:

ID X35012 standard; DNA; 1768 BP.

AC X35012; 1999 (first entry)

DE Genomic DNA sequence of PCV strain 999PCV.

KW PCV; strain PCV999; pig; PWS: porcine multisystemic wasting syndrome;

KW vaccine; ss.

OS Porcine circovirus

PN FR2769321-AL.

PD 09-APR-1999.

PF 03-OCT-1997; 012382.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PA (UYBE-) UNIV QUEENS BELFAST.

PA (UYSA-) UNIV SASKATCHEWAN.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246947/21.

PT New porcine circovirus from animals with porcine systemic wasting

PT Syndrome

PS Claim 13; Fig 1; 35pp; French.

CC The present sequence represents the genomic sequence of porcine
CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
CC isolated from a physiological or tissue sample, particularly from
CC a lesion, from a pig showing symptoms of PWS (porcine multisystemic
CC wasting syndrome), or cultured cells, infected with PCV isolated from
CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.

CC Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:

Quality: 681.00 Length: 128

Ratio: 5.405 Gaps: 0

Percent Similarity: 98.438 Percent Identity: 97.656

alignment_block:

US-09-209-961-5 x X35012/rev ..

Align seg 1/1 to reverse of: X35012 from: 1 to: 1768

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122

1768 TGGCCCTGCTCCCTCCATCACCAGGGGTAGGGAGTGGGCTCCACTGC 1719

122 aValIleLeuAspAspAsnPhaValThrLysAlaThrAlaLeuThrTyra 139

1718 TGTATTCTAGATGATACTTTGTACAAAGGCCACGCCCTAACCTATG 1669

139 spProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155

1668 AECATATGTAAACTACTCTCCCGGCATACATATCCCAACCTCTCTCC 1619

```
156 TyrHisserArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1618 TACACTCCCGTTACTTCACACCAACCTGTTCTTGACTCCACTATGA 1569
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1568 TTACTTCCACCAAAATAACAAAGGAATACGTTTGCTGAGGCTACAAA 1519
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1518 CCTCTAGAAATGTGACACGAGTAGCGCTCGCAGCTGCGTTTCGAAACAGT 1469
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1468 ATATACGACACGAGGTACAAATATCGGTGAACCATGTATGTACAATTCAG 1419
|||||
222 gGluPheAsnLeuLysAspProLeuGluPro 233
|||||
1418 AGAATTTAATCTTAAAGACCCCTTAAACCC 1385
```

seq_name: N_Geneseq_36:X35212

seq_documentation_block:

ID X35212 standard; DNA; 1768 BP.

AC X35212;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Imp999 (corrected version).

KW PCV isolate; type II porcine circovirus; PCV; PWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS Porcine circovirus.

PN FR2769322-AL.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows

PS Claim 14; Fig 3; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.

CC Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:

Quality: 681.00 Length: 128

Ratio: 5.405 Gaps: 0

Percent Similarity: 98.438 Percent Identity: 97.656

alignment_block:

US-09-209-961-5 x X35212/rev ..

Align seg 1/1 to reverse of: X35212 from: 1 to: 1768

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122

1767 TGGCCCTGCTCCCTCCATCACCAGGGGTAGGGAGTGGGCTCCACTGC 1718

122 aValIleLeuAspAsnPheValThrLysAlaThrLysAlaLeuThrTyra 139
 |||||
 1717 TGTATTCTAGATGATACTTTGTAAACAAAGCCACAGCCCTAACCTATG 1668
 139 sProTyrrValAsnTyrrSerArgHisThrIleProGlnProPheSer 155
 |||||
 1667 ACCCATATGTAACCTACTCTCCCGCATACAACTCCCAACCTCTCTCC 1618
 156 TyrrHisSerArgTyrrPheThrProLysProValLeuAspSerThrIleAs 172
 |||||
 1617 TACCACCTCCCGTTACTTACACCCAAACCTGTTCTTGACTCCACTATTGA 1568
 172 pTyrrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
 |||||
 1567 TTACTTCCAAACCAATAACAAAGGAATCAGCTTTGGCTGAGGCTACAAA 1518
 189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
 |||||
 1517 CCTCTAGAAATGTGGACACGCTAGGCTCGGCACCTGCTTGGAAACAGT 1468
 206 LysTyrrAspGlnAspTyrrAsnIleArgValThrMetTyrrValGlnPheAr 222
 |||||
 1467 ATATACGACGAGGACTACAAATATCCGTGTAAACCATGTATGTACAATTCAG 1418
 222 gGluPheAsnLeuLysAspProProLeuGluPro 233
 |||||
 1417 AGAATTAATCTTTAAAGACCCCTTAAACCC 1384

seq_name: N_Geneseq_36:X35214

seq_documentation_block:

ID X35214 standard; DNA; 1768 BP.

AC X35214;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS Porcine circovirus.

PN FR2769322-Al.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) Merial SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PT of pregnant sows

PS Claim 14; Fig 6; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Imp999. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.

SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:

Quality: 681.00

Ratio: 5.405

Percent Similarity: 98.438

Percent Identity: 97.656

alignment_block:

US-09-209-961-5 x X35214/rev ..

Align seq 1/1 to reverse of: X35214 from: 1 to: 1768

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
 |||||
 1768 TGGCCCTGCTCCCCCATCACCGGGTGATAGGGAGTGGGCTCCACTGC 1719
 122 aValIleLeuAspAsnPheValThrLysAlaThrLysAlaLeuThrTyra 139
 |||||
 1718 TGTATTCTAGATGATACTTTGTAAACAAAGCCACAGCCCTAACCTATG 1669
 139 sProTyrrValAsnTyrrSerArgHisThrIleProGlnProPheSer 155
 |||||
 1668 ACCCATATGTAACCTACTCTCCCGCATACAACTCCCAACCTCTCTCC 1619
 156 TyrrHisSerArgTyrrPheThrProLysProValLeuAspSerThrIleAs 172
 |||||
 1618 TACCACCTCCCGTTACTTACACCCAAACCTGTTCTTGACTCCACTATTGA 1569
 172 pTyrrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
 |||||
 1568 TTACTTCCAAACCAATAACAAAGGAATCAGCTTTGGCTGAGGCTACAAA 1519
 189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
 |||||
 1518 CCTCTAGAAATGTGGACACGCTAGGCTCGGCACCTGCTTGGAAACAGT 1469
 206 LysTyrrAspGlnAspTyrrAsnIleArgValThrMetTyrrValGlnPheAr 222
 |||||
 1468 ATATACGACGAGGACTACAAATATCCGTGTAAACCATGTATGTACAATTCAG 1419
 222 gGluPheAsnLeuLysAspProProLeuGluPro 233
 |||||
 1418 AGAATTAATCTTTAAAGACCCCTTAAACCC 1385

seq_name: N_Geneseq_36:X35210

seq_documentation_block:

ID X35210 standard; DNA; 1767 BP.

AC X35210;

DT 01-JUL-1999 (first entry)

DE Nucleotide sequence of PCV isolate Impl011-48121.

KW PCV isolate; type II porcine circovirus; PCV; PMWS;

KW porcine multisystemic wasting syndrome; pig; vaccine; ss.

OS Porcine circovirus.

PN FR2769322-Al.

PD 09-APR-1999.

PF 22-JAN-1998; 000873.

PR 03-OCT-1997; FR-012382.

PA (MERI-) Merial SAS.

PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,

PI Haines D, Harding J, Hassard L, Meehan B;

DR WPI: 99-246948/21.

PT New type II porcine circovirus, used for, e.g. passive immunization

PT of pregnant sows

PS Claim 14; Fig 1; 48pp; French.

CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl011-48121. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.

SQ Sequence 1767 BP; 447 A; 360 C; 502 G; 458 T;

alignment_scores:
 Quality: 669.00 Length: 128
 Ratio: 5.310 Gaps: 0
 Percent Similarity: 98.438 Percent Identity: 95.312

alignment_block:
 US-09-209-961-5 x X35210/rev ..

Align seg 1/1 to reverse of: X35210 from: 1 to: 1767

```

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1766 TGGCCCTGCTCCCGATCACCAGGCTGACAGGGAGTGGGCTCCAGTGC 1717
|||||
122 aValIleLeuAspAspAsnPhelValThrLysAlaThrAlaLeuThrTyra 139
|||||
1716 TGTATTCTAGATGATACTTTGTAAAGGCGCACAGCCCTCACCTATG 1667
|||||
139 sProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155
|||||
1666 ACCCTATGTAACACTCTCTCCCGCCATACCATACCCAGCCCTCTCC 1617
|||||
156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1616 TACCACCTCCGCTACTTTACCCCAAAAGGCGTGGCTGAGACTATTCGA 1567
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1566 TTACTTCCAAACCAACAAAGAAAGGAGTGGCTGAGACTACAAA 1517
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1516 CTACTGGAAATGTAGACCACTAGTGGCTCGGCTCGGCTCGGCTCGGCT 1467
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1466 ATATACGACCAAGGATACAAATATCCGTGTAAACCATGTATGTACATTCAG 1417
|||||
222 gGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1416 AGAATTTAATCTTAAAGACCCCTTAAACCTTAAACCT 1383
|||||
seq_name: N_Geneseq_36.X35211

seq_documentation_block:
ID X35211 standard; DNA; 1767 BP.
AC X35211;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl011-48285.
KW PCV isolate, type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 2; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl011-48285. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual

```

alignment_scores:
 Quality: 669.00 Length: 128
 Ratio: 5.310 Gaps: 0
 Percent Similarity: 98.438 Percent Identity: 95.312

alignment_block:
 US-09-209-961-5 x X35211/rev ..

Align seg 1/1 to reverse of: X35211 from: 1 to: 1767

```

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1766 TGGCCCTGCTCCCGATCACCAGGCTGACAGGGAGTGGGCTCCAGTGC 1717
|||||
122 aValIleLeuAspAspAsnPhelValThrLysAlaThrAlaLeuThrTyra 139
|||||
1716 TGTATTCTAGATGATACTTTGTAAAGGCGCACAGCCCTCACCTATG 1667
|||||
139 sProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155
|||||
1666 ACCCTATGTAACACTCTCTCCCGCCATACCATACCCAGCCCTCTCC 1617
|||||
156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1616 TACCACCTCCGCTACTTTACCCCAAAAGGCGTGGCTGAGACTATTCGA 1567
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1566 TTACTTCCAAACCAACAAAGAAAGGAGTGGCTGAGACTACAAA 1517
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1516 CTGCTGGAAATGTAGACCACTAGTGGCTCGGCTCGGCTCGGCTCGGCT 1467
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1466 ATATACGACCAAGGATACAAATATCCGTGTAAACCATGTATGTACATTCAG 1417
|||||
222 gGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1416 AGAATTTAATCTTAAAGACCCCTTAAACCTTAAACCT 1383
|||||
seq_name: N_Geneseq_36.X35211

seq_documentation_block:
ID X35211 standard; DNA; 1767 BP.
AC X35211;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Impl011-48285.
KW PCV isolate, type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 2; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl011-48285. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual

```

CC hybridization or amplification assays. These polypeptides may also be
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment_scores:
 Quality: 669.00 Length: 128
 Ratio: 5.310 Gaps: 0
 Percent Similarity: 98.438 Percent Identity: 95.312

alignment_block:
 US-09-209-961-5 x X35211/rev ..

Align seg 1/1 to reverse of: X35211 from: 1 to: 1767

```

106 TrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAl 122
|||||
1766 TGGCCCTGCTCCCGATCACCAGGCTGACAGGGAGTGGGCTCCAGTGC 1717
|||||
122 aValIleLeuAspAspAsnPhelValThrLysAlaThrAlaLeuThrTyra 139
|||||
1716 TGTATTCTAGATGATACTTTGTAAAGGCGCACAGCCCTCACCTATG 1667
|||||
139 sProTyrValAsnTyrSerArgHisThrIleProGlnProPheSer 155
|||||
1666 ACCCTATGTAACACTCTCTCCCGCCATACCATACCCAGCCCTCTCC 1617
|||||
156 TyrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAs 172
|||||
1616 TACCACCTCCGCTACTTTACCCCAAAAGGCGTGGCTGAGACTATTCGA 1567
|||||
172 pTyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgLeuGlnT 189
|||||
1566 TTACTTCCAAACCAACAAAGAAAGGAGTGGCTGAGACTACAAA 1517
|||||
189 hrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGluAsnSer 205
|||||
1516 CTACTGGAAATGTAGACCACTAGTGGCTCGGCTCGGCTCGGCTCGGCT 1467
|||||
206 LysTyrAspGlnAspTyrAsnIleArgValThrMetTyrValGlnPheAr 222
|||||
1466 ATATACGACCAAGGATACAAATATCCGTGTAAACCATGTATGTACATTCAG 1417
|||||
222 gGluPheAsnLeuLysAspProProLeuGluPro 233
|||||
1416 AGAATTTAATCTTAAAGACCCCTTAAACCTTAAACCT 1383
|||||
seq_name: N_Geneseq_36.X35013

seq_documentation_block:
ID X35013 standard; DNA; 1759 BP.
AC X35013;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence which has homology to PCV sequence.
KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
KW vaccine; ss.
OS Sus sp.
PN FR2769321-Al.
PD 09-APR-1999.
PF 03-OCT-1997; 012382.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PA (UYBE-) UNIV QUEENS BELFAST.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246947/21.
PT New porcine circovirus from animals with porcine systemic wasting
PT syndrome
PS Disclosure; Fig 2; 35pp; French.
CC The specification describes a genomic sequence of porcine

```

CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunopassay format. The present sequence appears in the
 CC specification.
 SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment_scores:
 Quality: 478.00 Length: 126
 Ratio: 4.268 Gaps: 0
 Percent Similarity: 88.889 Percent Identity: 69.841

alignment_block:
 US-09-209-961-5 x X35013 ..

Align seg 1/1 to: X35013 from: 1 to: 1759

104 GluPheTrpProCysSerProileThrGlnGlyAspArgGlyValGlySe 120
 |||||:||||| :|||||:|||||:|||||:|||||:|||||:|||||:|||||
 1 GAATTTTACCACGAGACCCCATCCCTTATCAAGAGGTGTGGTGC 50
 120 rThraValleleuAspAspAspPheValThrLysAlaThrAlaLeuT 137
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 51 CACTGTGTTATCTGGTGCACACTTTGTATCCCTCCACCACTGG 100
 137 hrTyAspProTyValAsnTySerArgHisThrileProGlnPro 153
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 101 CCTATGACCCCTATATTAACCTACTCTCCGCCACACCAAGGAGCC 150
 154 PheSerTyHisSerArgTyPheThrProLysProValLeuAspSerTh 170
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 151 TTTACCTACCACTCCAGTACTTCAACCCCAACCTGAGCTGGACCAAC 200
 170 rIleAspTyPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeuArgL 187
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 201 AATTGATTGGTTCACCCCAATATAAAGAAACCAAGCTGTGGCTCCATT 250
 187 euGlnThrSerGlyAsnValAspHisValGlyLeuGlyThrAlaPheGlu 203
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 251 TAAATACCCACCAACATTCGAGCACACAGGCTCGGTATCGCTCCAA 300
 204 AsnSerLysTyAspGlnAspTyAsnIleArgValThrMetTyValG1 220
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 301 ANTGCACCCAGCCCAAAATATGTGTAGGCTGACTATTATTAATGTA 350
 220 nPheArgGluPheAsnLeuLysAspPro 229
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 351 ATTCAGAGAATTATCTCTAAAGACCT 378

seq_name: N_Geneseq_36:Q90194

seq_documentation_block:
 ID Q90194 standard; cDNA; 1848 BP.
 AC Q90194.
 DT 03-NOV-1995 (first entry)
 DE A. niger glucose-oxidase gene.
 KW Glucose-oxidase; transgenic plant; disease resistance;
 KW crop improvement; ss.
 OS Aspergillus niger.
 FH Key Location/Qualifiers
 FT cds 16..1833
 FT /*tag= a

PN W09514784-A.
 PD 01-JUN-1995.
 PF 25-OCT-1994; U11837.
 PR 24-NOV-1993; US-161041.
 PA (MONS) MONSANTO CO.
 PI Lawrence EB, Levine EB, Shah DM;
 DR WPI; 95-206941/27.
 DP P-PSDB; R74388.
 PT DNA for expressing Aspergillus glucose oxidase in plant(s), - to
 PT increase resistance to bacterial and fungal pathogen(s)
 PS Claim 2; Page 20-23; 32pp; English.
 CC Total DNA isolated from A. niger ATCC 9029 was used as template for
 CC PCR isolation of a glucose-oxidase gene. The PCR product was cloned
 CC into pUC118; its sequence is given in Q90194. Vectors incorporating
 CC e.g. tissue-specific or pathogen-inducible promoters are used to
 CC express the gene in plants to improve disease resistance.
 SQ Sequence 1848 BP; 387 A; 592 C; 470 G; 399 T;

alignment_scores:
 Quality: 91.50 Length: 280
 Ratio: 0.863 Gaps: 19
 Percent Similarity: 37.857 Percent Identity: 24.286

alignment_block:
 US-09-209-961-5 x Q90194 ..

Align seg 1/1 to: Q90194 from: 1 to: 1848

4 ProArgArgArgTyArgArgArgHisArgProArgSerHisLeuG1 20
 |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
 939 CCTGGCGGGGCTCCGCTCTCTCTCCACATCTCGA..... 977
 20 yGlnIleLeuArgArg.....ProTrpLeuV 30
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 978 ...ATATTCGGTATCGGAATGAAGTCCATCTGGAGCCCTTGGTATC 1023
 30 aHis.....ProArgHis 34
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1024 GACACCGTCTGACCTGCCGCTTGAACCTGCAGGACGACCAC 1073
 35 ArgTyArg.....Tip...ArgArgLysAsnG1 43
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1074 CGCTACCGCTCCGCTCCGATCACCCTCTGCTGTGAGGACAGGACAGG 1123
 43 yIlePheAsnThrArgLeuSerArg...ThrPheGly...TyThrValLys 58
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1124 CGCTTGGTTCGCCACCTTCAACGAGACCTTTGGTACTATTCGAAAG 1173
 59 .ArgThrThrValThrThrProSerTrpAlaValaspMetMetArgPheL 75
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1174 GCACAGAGCTGCTCAACACCAAGCTGGAGCAGTGG..... 1209
 75 ysIleAspAspPheValProGlyGly.....GlyThrAsnLysile 89
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1210 ..GCCAAGAGCGCGCTGCCGCTGGGATTCACACACCCCGCTTG 1257
 90 SerIleProPheGluTyTyArg..... 97
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1258 CTCATCCAGTACGAGAACTACCGGAGTGTCAACCAACACGCTGC 1307
 98IleArgLysValLysValGluPheT 106
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1308 GTACTCGAACTCTTCTCGACACTCGCGAGTAGCCAGCTTGGATGT 1357
 106 rProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAla 132
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1358 GGGACCTTCTGCCCTTCAACCGAGGA.....TAGGTT 1389
 123 ValIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThrTyAs 139
 |||||:|||||:|||||:|||||:|||||:|||||:|||||
 1390 CACATCTCGAAG.....GA 1406

```

139 pProTyrValAsnTyrSerSerArgHisThrIleProGlnProPheSerT 156
      |||||:::
1407 CCCTACTCTTCCACCAC.....TTCCGCT 1429
      |||||

156 yrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAsp 172
      ||:::|||||
1430 ACGACCCCTCAGTACTTCTC..... 1449
      |||||

173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 195
      |||||:::||||
1450 .....AACGAGCTGCAGTCTCGCTCAGC 1475
      |||||

186 ...ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
      |||:::|||||:::||||:::
1476 TGCCTCTACTCACTGGCCCGCCACACATCTCCAACTCGGTGCCATGAGA 1525
      |||||

200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
      |||||
1536 CCTACTTCGCTGGGGAGACTATCCCGGTGATAACCTCGGTATGATGCC 1575
      |||||

210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
      |||||
1576 GATTTGGAC...GCTGGACTGAGTACATCCCGTAC 1608
      |||||

```

seq_name: N_Geneseq_36:T29485

```

seq_documentation_block:
ID   T29485 standard; DNA; 1848 BP.
AC   T29485;
DE   20-NOV-1996 (first entry)
DT   DE Aspergillus niger glucose oxidase gene.
KW   Glucose oxidase; expression vector; transgenic plant; resistance;
KW   fungal; bacterial; pathogen; potato; ds.
KW   Aspergillus niger.
FH   Key Location/Qualifiers
FT   16..1833
FT   cds /*tag= a
FT   PT
FN   US5516671-A.
PD   14-MAY-1996.
PF   24-NOV-1993; 161041.
PR   24-NOV-1993; US-161041.
PR   03-NOV-1994; US-333802.
PA   (MONS ) MONSANTO CO.
PI   Lawrence EB, Levine EB, Shah DM;
DR   WPI: 96-251010/25.
DR   P-PSTDB: R93i73.
PT   DNA mol. comprising promoter, Aspergillus glucose oxidase gene and
PT   polyA signal - used in prodn. of transgenic plants, esp. potato(s)
PT   resistant to bacterial and fungal pathogens
PS   Claim 2: Columns 13-18; l2pp: English.
CC   The present sequence, which encodes A. niger glucose oxidase (AGO),
CC   was used in the construction of a recombinant double stranded DNA
CC   mol., comprising a promoter functional in plants, the present
CC   sequence and a 3'-untranslated region, functional in plant cells,
CC   which adds polyA nucleotides to the 3'-end of the RNA sequence. The
CC   DNA mol. enables AGO to be expressed safely in plants, i.e.
CC   transgenic plants develop normally, which imparts resistance to
CC   fungal and bacterial pathogens (e.g. Verticillium dahliae,
CC   Phytophthora infestans, Botrytis cinerea, Septoria nodorum,
CC   Pseudocercospora herpotrichoides, Gaumannomyces graminis and
CC   Erwinia carotovora) in the plants, specifically potatoes.
CC   Potato tubers transformed with the DNA mol. were incubated with
CC   50000 c.f.v. E. carotovora for 3 days at 23 degrees C at high
CC   humidity. The final number of bacteria/lesion for tubers that
CC   expressed Ago at high level was 0.855-4.55 billion, compared to
CC   16.76 billion for controls.
SQ   Sequence 1848 BP: 387 A; 592 C; 470 G; 399 T;
```

alignment scores:

alignment_scores:		
Quality:	91.50	Length: 280
Ratio:	0.863	Gaps: 19
Percent Similarity:	37.857	Percent Identity: 24.286

```

alignment_block:
  US-09-209-961-5 x T29485  ..
  Align seg 1/1 to: T29485 from: 1 to: 1848

4  ProArgArgArgTyrArgArgArgArgHlsArgProArgSerHisLeuG1 20
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
939 CTTGGCCGCGGCTCGGTGCTCTCTCCCAATCTCGA..... 977
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
20 yGlnIleLeuArgArgArg.....ProTrpLeuV 30
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
978 ...ATATCCGGTATCGAATGAAGTCCATCTGGAGCCCTTGGTATC 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
30 aLHis.....ProArgHls 34
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1024 GACACCGCTGTTGACCTGCCGCTCGGCTTGAACCTGCAGGACGACACCAC 1073
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 ArgTyrArg.....Trp...ArgArgLysAsnG1 43
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1074 CGCTACCGCTCCGCTCCCGCATCCCTCTCTGTCGAGGACGAGGACAGG 1123
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
43 yIlePheAsnThrArgLeuSerArg_ThrPheGly...TyrThrVallys 58
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1124 CGCCTTGGTTCGCCACCTTCAACGAGACCTTTGGTGACTATTCCGAAAG 1173
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
59 _ArgThrThrValThrTrpSerTrpAlaValAspMetMetArgPheL 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1174 GCACACGAGCTGCTCAACACCAAGCTGGAGCAGTGG..... 1209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
75 ysIleAspAspPheValProProGlyGly.....GlyThrAsnLysIle 89
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1210 ..GCCGAGAGCCGCTGCCGCTGGCGGATTCACACACACCCGCGCTTG 1257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
90 SerIleProPheGluTyrArg.....IleArgLysValLysValGluPheT 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1258 CTCATCCAGTACGAGAAGCTACCCGCGACTGGATTGTCAACCAACAGTCGC 1307
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
98 .....IleArgLysValLysValGluPheT 106
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1308 GTACTCGGAACCTCTCTCGACACTGCCGGAGTAGCCAGCTTCGATGTGT 1357
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
106 rpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAla 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1358 GGGACCTCTCGCCTTACCCGAGGA.....TACGTT 1389
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
123 ValIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThrTyrAs 139
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1390 CACATCTCGACAAG.....GA 1406
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
139 pProTyrValAsnTyrSerSerArgHisThrIleProGlnProPheSerT 156
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1407 CCCCACCTTCCACCAC.....TTCGCT 1429
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
156 yrHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAsp 172
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1430 AGGACCTCAGTACTCTC..... 1449
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 185
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1450 .....AACGAGCTGGACCTGCTCGGTCCAGGC 1475
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
186 ....ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1476 TGGCGCTACTCAACTGGCCCGCAACATCTCCAACTCCGGTGGCATGCAGA 1525
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1526 CCTACTTCGCTGGGAGACTATCCCGGTGATAACCTCCGATGATGATGCC 1575
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1576 GATTGTGAGC...GCCTGGAGCTAGTACATCCCGTAC 1608
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


PS Claim 1; Page 74-75; 118pp; Japanese.
 CC The present sequence represents DNA which participates in the regulation
 CC of expression of PEBP2 alpha A gene. The DNA produces a regulator
 CC protein with the activity of promoting bone formation and can serve as a
 CC promoter for prevention and treatment of bone diseases including
 CC osteoporosis.
 SQ Sequence 1015 BP; 163 A; 376 C; 327 G; 149 T;

alignment_scores:
 Quality: 87.50 Length: 196
 Ratio: 1.029 Gaps: 11
 Percent Similarity: 43.367 Percent Identity: 25.510
 alignment_block:
 US-09-209-961-5 x X33246 ..
 Align seg 1/1 to: X33246 from: 1 to: 1015

4 ProArgArgTyrArgArgHis.....ArgProAr 16
 460 CCTCGGAGGAGCGGAGGAGGCGCGGCAAGCAGCGCGCGCG 509
 16 gserHisLeuGlyGlnIleLeuArgArgArgProTyrLeuValHisProA 33
 510 AGCT.....GGGCAG.....CCGC 523
 33 rgHisArgTyrArgTyrArg.....ArgLysAsn 42
 524 GCTCGTCTGGCGTGGCGGACGAGGAGGCGCACACGCGCGAGAT 573
 43 GlyIlePheAsnThrArgLeuSerArgThrPhe.....GlyTyrTh 56
 574 GGACTGTGAACCTCGCGGGTCTACTACCTGCTGGAACCGGAGCGG 623
 56 rValLysArgThrThrValThrPro..... 65
 624 CGCGCAGCGCACCGGACGCGGTGCCCGCAAGGGGCCCTACTGCAAGCT 673
 66SerTyrAlaValAspMetMetArgPhe 74
 674 GTTAACCTCAAGTCCTTCGCGGCTGGGCCAACACGCGCGCGTCC 723
 75 LysIle.....AspAspPheValProProGlyGlyGlyTh 86
 724 CGCTCGTAGCAGCCAGCCCGGACTCCCAAGCCTCCCGGTGGAGATCC 773
 86 rAspLysIleSerIleProPheGluTyrTyrArgIleArgLysValLysV 103
 774 CGGCTCTCGGGTGTCCCGGCTACCCGACTCCCGACTGTGCGGTCTCCAG 823
 103 alGluPheTyrProCysSerProIleThrGlnGlyAspArgGlyValGly 119
 824 TC.....GCGGCGCCCGCAGCAGACTCCGCGCGCGCACCGGCT 861
 120 SerThrAlaValIleLeuAspAspAsnThrValThrLysAlaThrAlaLe 136
 862 TCCCGAACTTGAT.....TCTACCTCTCTGCTCCCGCTACCTCCAT 905
 136 uThrTyrAspProTyrValAsnTyrSerSerArgHisThrIleProGlnP 153
 906 CCTCTTCGCGCC.....CGGCTCGGCTCCAC.....CCCTCG 940
 153 ropheSerTyrHisSerArgTyrPheThrProLysPro 165
 941 ATTTCCTCTCTCGCGCCCGCTATTCACCTCTCTCC 978

seq_name: N_Geneseq_36:062817

seq_documentation_block:
 ID 062817 standard: DNA; 1738 BP.
 AC 062817:
 DT_10-NQW-1994 (first entry)

DE Genomic sequence encoding human alpha 1 B adrenergic receptor.
 KW Alpha 1B; adrenergic receptor; antagonist; neuroreceptor; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 124..1683
 FT cds /*tag= a
 PN WO9410989-A.
 PD 26-MAY-1994.
 PF 12-NOV-1993; U10950.
 PR 13-NOV-1992; US-975867.
 PA (SYNA-) SYNAPTIC PHARM CORP.
 PI Branchek TA, Chiu G, Forray CC, Gluchowski C, Hartig PR;
 PI Wetzel JM;
 DR WPI: 94-183130/22.
 DR P-PSDB: R52831.
 PT Treating benign prostatic hyperplasia - with cpds. which bind
 PT selectively to alpha1C adrenergic receptor to inhibit contraction
 PT of prostatic tissue
 PS Example: Page 46-51; 69pp; English.
 CC The inventors claim a method of treating benign prostatic
 CC hyperplasia which comprises administering a therapeutically
 CC effective amount of a compound which binds to a human alpha-1C
 CC adrenergic receptor with a binding affinity greater than ten-fold
 CC higher than the binding affinity with which the compound binds to a
 CC human alpha-1A adrenergic receptor or a human alpha-1B adrenergic
 CC receptor. The example concerns a protocol for the determination of
 CC the potency of alpha-1 antagonists. Cell lines were prep. by
 CC transfecting the cloned cDNA or cloned genomic DNA or constructs
 CC contg. both genomic DNA and cDNA encoding human alpha-adrenergic
 CC receptors. For alpha-1A, the entire coding region and 5' and 3'
 CC UTRs were cloned into the BamHI and ClaI sites of the
 CC polylinker-modified eukaryotic expression vector pCEXV-3, called
 CC Exu.HR. For alpha-1B, the entire coding region and 5' and 3' UTRs
 CC were cloned into the EcoRI site of pCEXV-3 eukaryotic expression
 CC vector. For alpha-1C, the entire coding region and 5' and 3'
 CC UTRs were cloned into the KpnI site of the polylinker-modified
 CC pCEXV-3-derived eukaryotic expression vector, EXJ.RH.
 SQ Sequence 1738 BP; 308 A; 610 C; 495 G; 325 T;

alignment_scores:
 Quality: 90.00 Length: 127
 Ratio: 1.385 Gaps: 8
 Percent Similarity: 51.181 Percent Identity: 28.346
 alignment_block:
 US-09-209-961-5 x Q62817/rev ..
 Align seg 1/1 to reverse of: Q62817 from: 1 to: 1738
 4 ProArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20
 1268 CCCAGCGACGCGCGCGGCTCGCGCGCGCGCGCGCG..... 1227
 20 yGlnIleLeuArgArgProTyrLeuValHisProArgHisArgTyrA 37
 1226CGGCACTGG.....CACCCGAGGATCGCGCAGCA 1199
 37 rGtrpArgArgLysAsnGlyIlePheAsn.....ThrArgLeuSer 50
 1198 AACCGCGCTTGAAGTCCTTCGTGGAGCATGGGTAGATGGGTGAGG 1149
 51 ArgThrPheGlyTyrThrValLysArgThrThrValThrProSerTr 67
 1148 CAGCTGTTGAAGTAGCCAGCCAGAACACACCTTGAACACGCGCTCGG 1099
 67 palavalAspMetMetArgPheLysIleAspAspPheValProGly. 83
 1098 GGGC.....TTCAGGTTGGAGAACAAAGAGCCCAAGCGGTA 1064
 84GlyGlyThrAsnLysIleSerIlePro..... 92
 1063 GAGCGATGAAGAGGGTAGCCAGCACACAGATGNACATACCGACCACATG 1014

93PheGluTyrTyrArgIleArgLysVallys..... 102
 1013 CCAAGCTCTTAGCTTCTTTCTTCTGGAGAACTTAAAGATTGAC 964
 103ValGluPhe.TrpProCysSerPro 110
 963 AGCTATGGAAGTCTCTGGGTTGTGGCCT 935

seq_name: N_Geneseq_36:X20631

seq_documentation_block:

ID X20631 standard; DNA; 722 BP.
 AC X20631;
 DT 05-MAY-1999 (first entry)
 DE Polynucleotide sequence from the genome of Treponema pallidum.
 KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
 KW enzyme production; ds.
 OS Treponema pallidum.
 PN W09859034-A2.
 PD 30-DEC-1998.
 PF 23-JUN-1998; U13041.
 PR 24-JUN-1997; US-050667.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Fraser CM;
 DR WPI; 99-081273/07.
 PT New isolated Treponema pallidum nucleic acids - used to develop
 products for the detection, diagnosis, characterisation, prevention
 and therapy of T. pallidum infections, particularly syphilis
 PS Claim 1; Page 763; 1150pp; English.
 CC X20500-21243 represent polynucleotide sequences from the genome of
 Treponema pallidum. The sequences can be used for detection,
 diagnosis, characterisation, prevention and therapy for T. pallidum
 CC infections, particularly syphilis. They can also be used for detecting
 diseases related to Borrelia infections in animals, and for the
 CC production of biosynthetic products such as enzymes.
 SQ Sequence 722 BP; 183 A; 152 C; 221 G; 166 T;

alignment_scores:
 Quality: 83.50 Length: 187
 Ratio: 0.949 Gaps: 12
 Percent Similarity: 47.059 Percent Identity: 26.738
 alignment_block:
 US-09-209-961-5 x X20631 ..

Align seg 1/1 to: X20631 from: 1 to: 722
 1 MetThrTyrProArgArgTyrArgArgArgHisArgProArg.. 16
 27 TTGAGGTATTTGGCGGAAAGGTTCTCTCGAAGCAGG...AGACCGAAGAC 73
 17SerHisLeuGlyGlnIleLeuArgA 25
 74 ACAGCGGGAAGAAAAGTGGGAGTACGAGACGACCCAGCGTTACTAA 123
 25 rGArgProTrpLeuValHisProArgHisArgTyrArgTrpArgLys 41
 124 GA.....TGGTCGGTCCCTCGCGTCATTCAGGATTTGGGAGAGCAG 167
 42 AsnGlyIlePheAsnThrArgLeuSerA-gThrPheGlyTyrThrVally 58
 168 GGG.....AGATTAAAGT.....TTGAA 184
 58 sArgThrThrVal.....ThrThrProSerTrpA 68
 185 GCAGTCGAGGGTGCAGTAGCGTTGGCGGATCGCGGAGTTCTTCATGGT 233
 68 lavalAspMetMetArgPheLysIleAspAsp..... 78
 234 T....GACAGCGAGGANTACAGATTACGNACTAAGGTTTCAGGTATG 280

79 ...PheValProProGlyGlyGlyThrAsnLysIleSerIleProPheG1 94
 281 AAGTTTGTCCCA.....GTTGCGTTCCTCATGA 309
 94 uTyTyrArgIleArgLysValGluPheTrpProCysSerProI 111
 310 ATTAAGGGGATTCAGGAGGAGGTTTCTCTCGTGAAGACATCCCGC 359
 111 leThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspAsp 127
 360 TTACGGAGATACCAACGCCCTT.....AAGACAATGCTCCTCAGGAT 403
 128 AsnPheValThrLysAlaThrAlaLeuThrTyrAspProTyrValAsnTY 144
 404 AGTTTCTGCACTAAGGTAAGCAGCATGAGAGCCGCGACGA....CC 449
 144 rSerSerArgHis.....ThrIleProGlnProPheSerTyrHis 158
 450 TTGTGTAGACACGGTGGGTACCGGTTACACAGCCGTTTGTGTTCCGAC 499
 158 erArgTyrPhe 161
 500 GCAGAGGCTTC 510
 seq_name: N_Geneseq_36:X18977
 seq_documentation_block:
 ID X18977 standard; DNA; 3722 BP.
 AC X18977;
 DT 13-MAY-1999 (first entry)
 DE Non-B, non-C, non-G hepatitis virus gene DNA sequence SEQ ID NO:62.
 KW Hepatitis virus; non-B non-C non-G hepatitis virus; PCR primer;
 KW virion; detection; screening; infection; blood transfusion; ss.
 OS Hepatitis virus.
 PN W09905282-A1.
 PD 04-FEB-1999.
 PF 27-JUL-1998; J03340.
 PR 13-MAR-1998; JP-082962.
 PR 25-JUL-1997; JP-233246.
 PR 09-OCT-1997; JP-314196.
 PA (TAMU/) TAMURA R.
 PI Nishizawa T, Okamoto H;
 DR WPI; 99-142937/12.
 PT Gene isolated from non-B non-C non-G hepatitis DNA virus - and its
 expression products, useful for diagnosis and treatment of hepatitis
 PT and screening of blood for transfusion.
 PS Claim 37; Page 102-103; 113pp; Japanese.
 CC The present sequence represents a gene which is contained in a non-B
 CC non-C non-G hepatitis DNA virus. It is 3500-4000 bases in length and
 CC contains two overlapping open reading frames (ORF). It is obtained by
 CC polymerase chain reaction (PCR) amplification using primer. The gene can
 CC be used for the production of vaccines for prevention and treatment of
 CC non-B non-C non-G hepatitis infection. Diagnosis of such infection, and
 CC screening of blood (e.g. intended for transfusion) for the presence of
 CC the virus, by using the virion or antigenic peptides as reagents for
 CC detection of antibodies to the virus, or by direct detection of the gene
 CC using PCR with primers derived from the gene sequence.
 SQ Sequence 3722 BP; 1092 A; 1029 C; 825 G; 776 T;

alignment_scores:
 Quality: 91.50 Length: 232
 Ratio: 0.934 Gaps: 14
 Percent Similarity: 42.241 Percent Identity: 25.862
 alignment_block:
 US-09-209-961-5 x X18977 ..
 Align seg 1/1 to: X18977 from: 1 to: 3722
 4 ProArgArgArgTyr.....ArgArgArgHisArgProAr 16
 623 CCGAGCGCGCGTGGAGAGGTGCGCGCGCGTGAAGGAGACTACCGAG 672


```
1024 GACACCGTCGTTGACCTGCCGTCGCTTGAACCTGCAGGACCAACAC 1073
1025
35 ArgTyrArg.....Trrp...ArgArgLysAsnG1 43
1074 CGTACCGTCGCTGCCGTCATCACCTCTGCTGGTCAGGACGAGGACAG 1123
1075
43 yllePheAsnThrArgLeuSerArg, ThrPheGly...TyrThrValLys 58
1124 CCGCTTGGTTCGCGACCTTCAACGAGAGCTTGGTGACTATTCCGAAAG 1173
1125
59 .ArgThrThrValThrThrProSerTrrpAlaValAspMetMetArgPheL 75
1174 GCACAGAGCTCTCAACACCAAGCTGGAGCAGTGG..... 1209
1175
75 ysileAspAspPheValProProGlyGly.....GlyThrAsnLysile 89
1210 .GCCGAAGAGCGCTGCCGCTGCCGCTGGCGGATTCACCAACACCAAC 1257
1211
90 SerileProPheGluTyrTyrArg..... 97
1258 CTCATCCAGTACGAGAACTACCGCGCTGGAGTGTCAACCAACACGTCGC 1307
1259
98 .....IleArgLysValLysValGluPheT 106
1308 GTACTCGGAACCTTCCTCGACACTGCCGAGTACCGACTTCGATGTGT 1357
1309
106 rpProCysSerProIleThrGlnGlyAspArgGlyValGlySerThrAla 122
1358 GGGACCTTCTGCCCTTACCGCGAGA.....TACGCT 1389
1359
123 valIleLeuAspAspAspPheValThrLysAlaThrAlaLeuThrTyrAs 139
1390 CACATCTCTGACAG..... 1406
1391
139 pProTyrValAsnTyrSerArgHisThrIleProGlnProPheSerT 156
1407 CCCCCTACCTTACCCAC.....TTCGCT 1429
1408
156 yRHisSerArgTyrPheThrProLysProValLeuAspSerThrIleAsp 172
1430 ACACCCCTCAGTACTCTC..... 1449
1431
173 TyrPheGlnProAsnAsnLysArgAsnGlnLeuTrpLeu..... 185
1450 .....AACGAGCTGGACCTGCTCGGTACAGC 1475
1451
186 ....ArgLeuGlnThrSerGlyAsnValAspHisValGly...LeuGlyT 200
1476 TGGCGCTACTCAACTGCCGCGCAACATCTCCAACTCCGGTGCATGCAG 1525
1477
200 hrAlaPhe.....GluAsnSerLysTyrAspGln 209
1526 CCTACTTCGCTGGGAGACTATCCCGCGTGATACCTCGCGATGATGCC 1575
1527
210 AspTyrAsnIleArgValThrMetTyrValGlnPhe 221
1576 GATTGAGC...GCCTGGAGTACGATACCTCCGCTAC 1608
```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq.us-08-307-499-53

```
seq_documentation_block:
; Sequence 53, Application US/08307499
; Patent No. 5651972
; GENERAL INFORMATION:
; APPLICANT: Moyer, Richard W.
; APPLICANT: Vi ueta, Eladio
; APPLICANT: Gibbs, E.P.J.
; TITLE OF INVENTION: Use of Recombinant Swine Poxvirus as a
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: David R. Saliwanchik
```

```
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: U.S.A.
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/307.499
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,241
FILING DATE: 1-JUL-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/908,630
FILING DATE: 29-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/342,212
FILING DATE: 21-APR-1992
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: UF35.1.FWCC1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 904-375-8100
TELEFAX: 904-372-5800
INFORMATION FOR SEQ ID NO: 53:
SEQUENCE CHARACTERISTICS:
LENGTH: 1023 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
US-08-307-499-53
```

alignment_scores:
Quality: 85.00 Length: 126
Ratio: 1.545 Gaps: 6
Percent Similarity: 43.651 Percent Identity: 27.778

alignment_block:

US-09-209-961-5 x US-08-307-499-53

Align seg 1/1 to: US-08-307-499-53 from: 1 to: 1023

```
38 TrrpArgArg.....LysAsnGlyllePheAsnThrArgLeuSerAr 51
418 TGGACACGATTCGACCCCAAGATAAACCTTAGTAAGTTAGATCAGC 467
51 gThrPheGlyTyrThrValLysArgThrThrValThrProSerTrrpA 68
468 TACATTTCCA...AAGGTAGGAATGTAACTGTAAATCACAACACTGGG 514
68 laValAspMetMetArgPheLysIleAspAspPheValProGlyGly 84
515 AACGTGTTATGAACATCTT.....GGA 537
85 GlyThrAsnLysIleSerIleProPhe.....GluTyrTyrArgI 98
538 TCATGGGTAAACATTACATGTCCTGTTACAGCGGATGATATATAAAT 587
98 eArgLysValLysValGlu.....PheTrrpProCysSerProIleThrG 113
588 TTCTAAGGGATATATAGATAAGCCAGTTAAGCTTACTGTTACAGTTACAG 637
113 lngLysArgGlyValGlySerThrAlaValIleLeuAspAspAsnPhe 129
```

```

:::||||| :::::|||| :::::|||||
638 GAATTGAAGAGGAGATATACTACATTGATATGCACATTTGTATAATCAT 687
130 ValThrLysAlaThrAlaLeu..... 136
:::||||| :::::|||||
688 TATCCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 737
137 .....ThrTyrAspProTyrValAsn 143
|||||
738 GGACATATCGTTATGATCCGTCAGTAAAT 765

```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-334-698-3

```

seq_documentation_block:
; Sequence 3, Application US/08334698
; Patent No. 5556753
; GENERAL INFORMATION:
; APPLICANT: Jonathan A. Bard et al.
; TITLE OF INVENTION: DNA Encoding Human Alpha 1 Adrenergic
; TITLE OF INVENTION: Receptors and Uses Thereof
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/334,698
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/952,798
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 376901
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: N
; ANTI-SENSE: N
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 124..1683
; OTHER INFORMATION:
US-08-334-698-3

```

alignment_scores: Quality: 87.00 Length: 128
Ratio: 1.318 Gaps: 9
Percent Similarity: 51.562 Percent Identity: 28.125

alignment_block:

US-09-209-961-5 x US-08-334-698-3/rev ..

Align seg 1/1 to reverse of: US-08-334-698-3 from: 1 to: 1738

```

4 ProArgArgArgTyrArgArgArgHisArgProArgSerHisLeuG1 20
|||||
1268 CCCAGCGACGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG..... 1227
20 yGlnIleLeuArgArgArgProTyrLeuValHisProArgHisArgTyrA 37
||| |||
1226 .....CGGCACCTGG.....CACCCGAGGATGCGCAGCA 1199
37 rgTyrArgArgLysAsnGlyIlePheAsn.....ThrArgLeu 49
||| |||
1198 AA.....GGCGCTTGAACCTCTGCTGGAGCATGGGTAGATGATGGGGTTG 1152
50 SerArgThrPheGlyTyrThrValLysArgThrThrValThrThrProSe 66
::: :::::
1151 AGCAGCTGTTGAAGTAGCCCGCAGCAGACACCACTTGAACACGGCGTC 1102
66 rTTPAlaValAspMetMetArgPheLysIleAspPheValProProG 83
| :::::
1101 GGGGGG.....TTCAGGGTGGAGAACAAAGGAGCAAGCG 1067
83 ly.....GlyGlyThrAsnLysIleSerIlePro..... 92
||
1066 GTAGAGCGCATGAAGAGGTAGCCAGCAGCAAGATGAACATACCGACCACA 1017
93 .....PheGlyTyrThrArgIleArgLysValLys..... 102
1016 ATGCCCAACGCTTAGCTGCTTCTTTCTCTGAGAGAACTTAAAGTTT 967
103 .....ValGluPhe.TyrProCysSerPro 110
::: |||
966 GACAGCTATGAAGTCTCTGGGTGTGGCCCT 935

```

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-228-932-3

```

seq_documentation_block:
; Sequence 3, Application US/08228932
; Patent No. 5578611
; GENERAL INFORMATION:
; APPLICANT: Charles Gluchowski, Carlos C. Forray, George Chiu,
; APPLICANT: Theresa A. Brancheck, John M. Wetzel and Paul R. Hartig
; TITLE OF INVENTION: USE OF ALPHA-1C SPECIFIC COMPOUNDS TO TREAT BENIGN
; TITLE OF INVENTION: PROSTATIC HYPERPLASIA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10112
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/228,932
; FILING DATE: 13-APR-1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41878-B/JPW/TEP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 977-9550
; TELEFAX: (212) 664-0525
; TELEX: (212) 422523 COOP UI
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1738 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```



```

66 rTrplavalasPmetMetarPhelylsleAspApheValprProg 83
| : : |
1101 GGGGCGC.....TTCAGGTGGAGAACAAGGACCAAGCG 1067

83 ly.....GlyGlyThrAsnLysIleSerIlePro..... 92
|| : : : : : || : : : : : || : : : : : || : : : : : ||
1066 GTAGAGCGATGAAGAAGGGTAGCCACACAAAGATGAACATACCACGCCACA 1017

93 .....PheGluTyTYrArgIleArgLysVallys..... 102
|||||: : : : : || : : : : : || : : : : : || : : : : : ||
1016 ATGCCCAACGCTTAGCTGCTTTCTTTTTCCCTGGAGAACTAAAAAGTTT 967

103 .....ValGluPhe.TrpProCySerPro 110
: : : : : || : : : : : || : : : : : || : : : : : ||
966 GACACGTATGGAACTCCTGGGGTGTTGGCCCT 935

seq_name: /cgn2_6/ptodata/1/1na/PCTUS9_COMB.seq: PCT-US95-04203-3
seq_documentation_block:
; Sequence 3, Application PC/TUS9504203
; GENERAL INFORMATION:
; APPLICANT: Charles Gluchowski, Carlos C. Forray, George
; APPLICANT: Chiu, Theresa A. Brancheke, John M. Wetzel and P
; TITLE OF INVENTION: USE OF ALPHA-1C SPECIFIC COMPOUNDS TO
; TITLE OF INVENTION: TREAT BENIGN PROSTATIC HYPERPLASIA
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: COOPER & DUNHAM LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.24
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/04203
; FILING DATE: 4-APR-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: White, John P.
; REGISTRATION NUMBER: 28,678
; REFERENCE/DOCKET NUMBER: 41878-D-PCT/JPW/AGL
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 278-0400
; TELEFAX: (212) 391-0525
; TELEX:

```

```

,
, INFORMATION FOR SEQ ID NO: 3:
, SEQUENCE CHARACTERISTICS:
, LENGTH: 1738 base pairs
, TYPE: nucleic acid
, STRANDEDNESS: single
, TOPOLOGY: unknown
, MOLECULE TYPE: DNA (genomic)
, HYPOTHETICAL: N
, ANTI-SENSE: N
, FEATURE:
, NAME/KEY: CDS
, LOCATION: 124..1683
, OTHER INFORMATION:
, PCT-US95-04203-3
,
, alignment_scores:
, Quality: 87.00
, Length: 128
, Ratio: 1.318
, Gaps:
, Percent Similarity: 51.562
, Percent Identity: 28.125
,
, alignment_block:
, US-09-209-961-5 x PCT-US95-04203-3/rev
,
,

```

```

; TELEFAX: 703-391-9035
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "synthetic DNA"
US-07-814-220-23

alignment_scores:
    Quality: 67.00      Length: 24
    Ratio: 3.350        Gaps: 0
    Percent Similarity: 83.333    Percent Identity: 54.167

alignment_block:
US-09-209-961-5 x US-07-814-220-23/rev ..
Align seg 1/1 to reverse of: US-07-814-220-23 from: 1 to: 72

      4  PROARGARGTYYARGARGARGARGHISARGPROARGSERHISLEUG1 20
          |||:::|||::|||::|||::|||::|||::|||::|||::|||
      72  CGCGACGCGCGCTCGCGGAGTACTGCGGCGACCCGACGACGCGTGC 23

      20  YGLNILEUARGARGARGPRO 27
          |||:::|||::|||::|||::|||::|||::|||::|||
      22  GCAGCTACTGCGGCGACGCGCG 1

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-07-812-421-23

seq_documentation_block:
; Sequence 23, Application US/07812421
; Patent No. 5932897
; GENERAL INFORMATION:
; APPLICANT: Caceci, Thomas
; APPLICANT: Toth, Thomas E.
; APPLICANT: Szumanski, Maria B.W.
; TITLE OF INVENTION: SYNTHETIC ANTIFREEZE PEPTIDE AND
; TITLE OF INVENTION: SYNTHETIC GENE CODING FOR ITS PRODUCTION
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WHITHAM
; STREET: Reston Intl. Center, 11800 Sunrise Valley Dr.,
; STREET: Suite 900
; CITY: Reston
; STATE: VA
; COUNTRY: USA
; ZIP: 20191
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/812,421
; FILING DATE: 23-DEC-1991
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/588,437
; FILING DATE: 25-SEP-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Whitham, Michael E.
; REGISTRATION NUMBER: 32,635
; REFERENCE/DOCKET NUMBER: CIT.016
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-391-2510
; TELEFAX: 703-391-9035
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 72 base pairs
; TYPE: nucleic acid

```



```
/ ANTI-SENSE: NO
/ ORIGINAL SOURCE:
/ ORGANISM: S.pristinaespiralis
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 1..474
/ OTHER INFORMATION: /product= "Partie du gene Snbd"
US-08-403-852D-14

alignment_scores:
  Quality: 75.50      Length: 82
  Ratio: 1.641        Gaps: 3
  Percent Similarity: 56.098      Percent Identity: 32.927

alignment_block:
US-09-209-961-5 x US-08-403-852D-14 ..
Align seg 1/1 to: US-08-403-852D-14 from: 1 to: 474

3 TyrProArgArgTyrArgArgArgHisArgProArgSerHisLe 19
: ::::::::::::::: ||||| ||| ::::
216 CATCCCGTCGCGAGCCCGTCGCGCGCCGACCGACGCGCTCGACG 265
: ::::::::::::::: ||||| ||| ::::
19 u...GlyGlnIleLeuArgArgProTrpLeuValHisProArgHisA 35
: ||| ::::::::::: ||||| |||
266 ACCTGGTGGGCTTCGTCAA.....CACCTCGTCTG 300
: ||| ::::::::::: ||||| |||
35 rgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArg 51
: :::: ||||| :::: ||||| |||||
301 CGCACCGACACCTCCGCGACCCACCTCCGCGGAACCTCGTCGACGCGT 350
: :::: ||||| :::: ||||| |||||
52 ThrPheGlyTyrThrValLysArgThrValThrProSerTrpAl 68
: :::: ||||| :::: ||||| |||||
351 GCG.....GCAGTTCGACCTCGCGCCTACACGACCGACGACATGC 391
: :::: ||||| :::: ||||| |||||
68 aValAspMetMetArgPheLysIleAspAspPheValProProGly 83
: ||| :::: ||| :::: |||||
392 CGTTCGAAAGCTCGTCGAGAGGTCAACCCGAGCGCTCCCTGGC 437
: :::: ||||| :::: ||||| |||||

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-910-973-21

seq_documentation_block:
/ Sequence 21 Application US/08910973
/ Patent No. 5795723
/ GENERAL INFORMATION:
/ APPLICANT: Tapscott, Stephen J.
/ TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
/ NUMBER OF SEQUENCES: 24
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
/ STREET: 1420 Fifth Avenue, Suite 2800
/ CITY: Seattle
/ STATE: WA
/ COUNTRY: USA
/ ZIP: 98101-2347
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/910,973
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 08/239,238
/ FILING DATE: 06-MAY-1994
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: WO PCT/US95/05741
/ FILING DATE: 08-MAY-1995
/ PRIOR APPLICATION DATA:
```

```
/ APPLICATION NUMBER: PCT/US96/17532
/ FILING DATE: 30-October-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Sheiness, Diana K.
/ REGISTRATION NUMBER: 35,356
/ REFERENCE/DOCKET NUMBER: FHC-1-10958
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 206-682-8100; 206-224-0735 (direct)
/ TELEFAX: 206-225-0779
/ INFORMATION FOR SEQ ID NO: 21:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1333 base pairs
/ TYPE: nucleic acid
/ STRANDEDNESS: double
/ TOPOLOGY: linear
/ MOLECULE TYPE: DNA (genomic)
/ ORIGINAL SOURCE:
/ ORGANISM: Mus musculus
/ IMMEDIATE SOURCE:
/ CLONE: neuroD3
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: 101..835
US-08-910-973-21

alignment_scores:
  Quality: 80.50      Length: 212
  Ratio: 0.847        Gaps: 8
  Percent Similarity: 44.811      Percent Identity: 22.170

alignment_block:
US-09-209-961-5 x US-08-910-973-21 ..
Align seg 1/1 to: US-08-910-973-21 from: 1 to: 1333

6 ArgArgTyrArgArgArgHisArgProArgSerHis.....LeuG1 20
: ||||| ||||| ||| ||| |||||
314 CGCGGAGGCGGCGAGGTGCGCTCGGTCGGGTCCGAGGCTCTGTGCA 363
: :::: ||| :::: ||||| |||
20 yGlnIleLeuArgArgProTrpLeuValHisProArgHisArgTyrA 37
: :::: ||| :::: ||||| |||
364 CTCCTCGCGGAGGAGTGTGCGTCGTCAAAGCCAGCATCGGAGCGCAAC 413
: :::: ||| :::: ||||| |||
37 rgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer.....Arg 51
: || :::: ||||| :::: |||||
414 GCATGCACAACTCAAGCTGCGCTGGACGCTTGCAGCGTGTCTGCC 463
: :::: ||| :::: ||||| |||
52 ThrPheGlyTyrThrValLysArgThrValThrProSerTrpAl 68
: :::: ||||| :::: ||||| |||
464 TCGTTCGCCGACGACCAAGCTCACCAGATTGAGACGCTGCGCTTCGC 513
: :::: ||| :::: ||||| |||
68 aValAspMetMet.....ArgPheLysIleAspAspPhe 80
: | :::: |||
514 CTCAACTACATCTGGGCCCTGGCTGAGACACTGGCGCTGGCAGATCAAG 563
: :::: |||
80 alProProGlyGlyThrAsnLysIleSerIleProPheGluTyrTyr 96
: ||||| :::: ||||| |||
564 GGCTCCCGGGGCGAGTGCCTCGGAGCGCTCTGCTCGCTCGCAG..... 607
: :::: ||||| :::: ||||| |||
97 ArgIleArgLysValLysValGluPheTrpProCys.SerProIleThrG 113
: ||| |||||
608 .....TGTCCTCCCTGTCTGCC 624
: :::: |||||
113 lnglyAspArgGlyValGlySerThr..... 121
: ||| |||||
625 CGGGCCCCCGAGCCCGGCGACGACACTGAGTCTGGGGTTCGGGGCCG 674
: :::: |||||
122 .....AlaValIleLeuAspAsnPheValThr..... 131
: ||| :::: |||||
675 CTGCTCCCTCCCTGCGCCACTGTGTCATCACCACCTCTGTGACCCAGTAGT 724
: :::: |||
132 .....LysAlaThrAlaLeuThrTyrAspP 140
```

725 CCCTCGGCTTCAGAAAGACTTACCTATGCGCGGGATCCCTTTTC 774
140 rGlyValAsnTyrSerSerArgHisThrIleProGlnProPheSerTyr 156
775 CTTTCTCGCTGCCAAGAGCTCTGTCACACAGACGCGCTTTTCATCC 824
157 HisSerArgTyrPheThrProLysProValLeuAspSerThrIleAspTyr 173
825 CAT.....ACCACAGGCTTTGTAAAGCAACATCAATACATT 862
173 rPheGlnProAsnAsnLysArgAsnGlnLeuTyr 184
863 CTTCTCCCTCCAGCTAAGAGCAATAATAGATGG 896

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-910-973-12

seq_documentation_block:
; Sequence 12, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910, 973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239, 238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHC-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; TYPE: nucleic acid
; LENGTH: 1268 base pairs
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; CLONE: 20A1 (neuroD3)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..768
; US-08-910-973-12

alignment_scores:
Quality: 80.00 Length: 188
Ratio: 0.909 Gaps: 9
Percent Similarity: 46.809 Percent Identity: 25.000
alignment_block:
US-09-209-961-5 x US-08-910-973-12
Align seg 1/1 to: US-08-910-973-12 from: 1 to: 1268
6 ArgArgTyrArgArgArgHisArgProArgSerHis.....LeuG1 20
265 AGGCGCGCGCGCGCGCGGCTCAAGGCCAACGATCGGAGCGCTGCTGCA 314
20 yGlnIleLeuArgArgProTyrIleValHisProArgHisArgTyrA 37
315 CTCGCTGCGCAGGAGCGCGCGCTCAAGGCCAACGATCGGAGCGCAAC 364
37 rGTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSer.....Arg 51
365 GCATGCACAACTTGAACGCGGCTGGACGCTGCGACGCTGCTGCC 414
52 ThrPheGlyTyrThrValLysArgThrThrValThrProSerTyrP 68
415 TCGTTCCCGCAGCACACCAAGCTCACCANAATCGAGACGCTGGCTTC 464
68 aValAspMetMet.....ArgPheLysIleAspAspPheV 80
465 CTACAACTACATCTGGGCTCTGCGCGGAGACACTGGCGCTGGCGATCA 514
80 alProProGlyGlyThrAsnLysIleSerIleProPheGluTyrTyr 96
515 GGCTGCGCGAGCGGTGCC..... 534
97 ArgIleArgLysValLysValGlnPheTyrProCys..... 108
535 CGGAGCGCTCTCGCGCGCAGTGGCTGCTCCCTGCTCCCGGTCCCC 584
109 SerProIleThrGlnGlyAsp.....ArgGlyValGlySerThrA 122
585 AAGCGCGCGCAGCAGCGGAGTCTCGGGGCTCAGGTGCGCGCGCGCT 634
122 laValIleLeuAspAsnPheValThrLysAlaThrAlaLeuThrTyr 138
635 CCGCGCTCTCTGACCCAGTAGCCAGCGCTCGGAGACTTCACCTAC 684
139AspProTyrValAsnTyrSerSer.....Ar 147
685 CGCGCGCGCGACCTGTTTCTCTCCCAAGCGCTGCCCAAGACTTGCT 734
147 gHisThrIleProGlnProPheSerTyrHis.....Sera 159
735 CCACACACAGCGCTCTTTCATTCTTACCACCTAGGCCCTTTTAGACA 784
159 rGlyTyrPheThrPro 163
785 GTTACTTTCCTCCCT 798
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-314-309A-5
seq_documentation_block:
; Sequence 5, Application US/08314309A
; Patent No. 5677141
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: FUKAGAWA, MASAO
; APPLICANT: IWAMI, MORITA
; APPLICANT: ARAMORI, ICHIRO
; APPLICANT: KOJO, HITOSHI
; TITLE OF INVENTION: PROCESS FOR PRODUCING 7-AMINOCEPHEM
; TITLE OF INVENTION: COMPOUND OR SALTS THEREOF
; NUMBER OF SEQUENCES: 34

OM of: US-09-209-961-5 to: EST:* out_format : pfs

Date: Dec 23, 1999 5:42 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

-MODEL=framet.p2n.model -DEV=xlp
-O=cn2.1/USPTO.spool/US09209961/runat.22121999.101043.17901/app_query.fasta.1
-DB=EST -QFW=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.500
-MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=score -ALIGN=15 -MODE=LOCAL
-OUTFWT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY
-WAIT -THREADS=1

Search information block:

Query: US-09-209-961-5
Query length: 233
Database: EST:*
Database sequences: 4089388
Database length: 1713965092
Search time (sec): 1572.090000

score_list:

Sequence	Strd Orig	ZScore	Escore	Len	Documentation
gb_est32:AU070895	+	93.00	152.57	676	! AU070895 AU070895 Rice cDNA fro
gb_gss31:A0638908	+	86.00	147.84	316	! A0638908 927P1-14H7.TP 927P1.T
gb_est31:A1673888	-	88.50	145.35	630	! A1673888 605038F09.xl 605 - End
gb_est36:A1895135	+	87.50	144.08	600	! A1895135 EST264578 tomato callu
gb_est37:A1460429	+	85.00	144.06	387	! A1460429 sb1d12.yl Gm-cl004 GI
gb_est32:A1736822	+	85.00	141.64	36.09	! A1736822 sb34a05.yl Gm-cl012 GI
gb_gss8:A0577736	+	87.00	141.39	719	! A0577736 nbxb0091H09r CUGR Rice
gb_est37:A1946727	+	86.00	141.07	623	! A1946727 bs30a08.yl Drosophila
gb_gss2:A1900445	+	82.50	141.04	38.98	! A1900445 sc05h05.yl Gm-cl012 GI
gb_gss2:CN5017H1	-	88.00	140.75	408	! A1108048 Drosophila melanogaste
gb_est36:A1392143	-	84.00	139.76	459	! A1392143 NCC2D977 Conidial Neur
gb_est32:A069671	+	81.50	139.46	47.73	! A069671 AU069671 Rice callus C
gb_gss4:A0175288	-	83.00	138.93	455	! A0175288 HS_3210.AL.G12_MR CTF
gb_est27:A1441014	+	82.00	138.71	52.60	! A1441014 sb63g11.yl Gm-cl004 GI
gb_est37:A1941495	+	82.00	138.60	33.30	! A1941495 sb90g09.yl Gm-cl017 GI
gb_est32:A1735906	+	82.00	138.43	54.51	! A1735906 sb20d09.yl Gm-cl007 GI
gb_est37:A1416813	+	82.00	138.38	54.86	! A1416813 sb18c05.yl Gm-cl005 GI
gb_est28:A1495208	+	82.00	138.38	54.86	! A1495208 sb02a08.yl Gm-cl004 GI
gb_est27:A1444041	+	82.00	138.35	55.03	! A1444041 sb31e11.yl Gm-cl004 GI
gb_est28:A1460767	+	82.00	138.35	55.03	! A1460767 sb69a05.yl Gm-cl004 GI
gb_est27:A1441174	+	81.50	138.24	55.81	! A1441174 sb52c03.yl Gm-cl004 GI
gb_est28:A1495366	+	82.00	138.21	56.08	! A1495366 sb97b04.yl Gm-cl004 GI
gb_est27:A1441122	+	82.00	138.13	56.61	! A1441122 sb59a12.yl Gm-cl004 GI
gb_est32:A1736457	+	82.00	137.87	58.54	! A1736457 sb28h05.yl Gm-cl009 GI
gb_est36:A1900949	+	82.00	137.75	59.43	! A1900949 sc17e10.yl Gm-cl013 GI
gb_est28:A1461214	+	82.00	137.73	59.61	! A1461214 sb76f05.yl Gm-cl004 GI
gb_est27:A1460639	+	82.00	137.71	59.79	! A1460639 sb71d09.yl Gm-cl004 GI
gb_est28:A1416528	+	82.00	137.62	60.50	! A1416528 sb10c04.yl Gm-cl003 GI
gb_est28:A1495104	+	82.00	137.50	61.39	! A1495104 sb03e02.yl Gm-cl004 GI
gb_est37:A1941368	+	85.00	136.97	65.68	! A1941368 sc12c06.yl Gm-cl013 GI
gb_est27:A1442364	+	82.00	136.95	65.89	! A1442364 sb26a06.yl Gm-cl004 GI
gb_est37:A1437962	+	82.00	136.55	69.35	! A1437962 sb41h09.yl Gm-cl004 GI
gb_est36:A1900279	+	82.00	136.47	70.19	! A1900279 sc03d06.yl Gm-cl012 GI
gb_est19:AA751860	-	82.00	136.25	72.12	! A1900279 sc03d06.yl Gm-cl012 GI
gb_gss10:A0641744	-	82.00	136.07	73.79	! A0641744 RPI193-DpnII-26J16.TJ
gb_est36:A1888176	+	83.50	135.93	75.12	! A1888176 sb90f05.xl NCI-CCAP.BH
gb_est1:D22673	+	79.50	135.85	75.82	! D22673 R1C0842A Rice callus O
gb_est32:A1727474	-	83.00	135.63	78.04	! A1727474 BNLGH18050 Six-day Cot
gb_est36:A1881562	-	81.50	134.95	85.11	! A1881562 606070G08.yl 606 - Ear
gb_est28:A1494957	+	78.00	134.73	87.59	! A1494957 sb93b05.yl Gm-cl004 GI
gb_est18:AA680493	+	81.50	134.56	89.53	! AA680493 SWAMCA3090SK Bruglia m
gb_est32:A0069559	+	80.00	134.40	91.38	! A0069559 AU069559 Rice callus d

gb_est31:A1692946 + 81.00 134.21 93.63 513 ! A1692946 wd89f08.xl NCI_CGAP
gb_est36:A1881449 - 81.50 134.16 94.22 563 ! A1881449 606069810.yl 606 -
gb_est5:N31093 - 80.50 134.01 96.00 479 ! N31093 yx63h03.rl Soares mel

seq_name: gb_est32:AU070895

seq_documentation_block:
LOCUS AU070895 676 bp mRNA EST 10-JUN-1999
DEFINITION AU070895 Rice cDNA from young root Oryza sativa CDNA clone
R10430.1A, mRNA sequence.

ACCESSION AU070895
NID 95038785
VERSION AU070895.1 GI:5038785
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.

REFERENCE 1 (bases 1 to 676)
AUTHORS Yamamoto K. and Sasaki, T.
TITLE Rice cDNA from young root
JOURNAL Unpublished (1999)
COMMENT On Jun 22, 1998 this sequence version replaced gi:3246692.

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abr.affrc.go.jp
PROJECT = "RGP".

FEATURES

source
1..676
/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="R10430.1A"
/clone_lib="Rice cDNA from young root"
/tissue_type="young root"
BASE COUNT 130 a 225 c 211 g 105 t 5 others
ORIGIN

alignment_scores:

Quality: 93.00 Length: 196
Ratio: 1.208 Gaps: 13
Percent Similarity: 39.286 Percent Identity: 26.531

alignment_block:

US-09-209-961-5 x AU070895
Align seg 1/1 to: AU070895 from: 1 to: 676
4 ProArgArgTyrArgArgArg..... 11
|||||
90 CCTCGGGCAGACACACCGCGGAGCTACCTGACACACGCGCG 139
12ArgHisArgPro..... 15
|||||
140 CGCGGGGACGCGCGCGGAGCTAGCTGAGTCCCAAGTCCGCGC 189
16 ..ArgSerHisLeu...GlyGlnIleLeuArgArgProTrpLeuVal 30
|||||
190 TTCGCGAGAGCTACGCGGCGCAAGCGCGGAGCTGCGGCTGNCAGC 239
31 HisProArgHisArg..... 35
240 ACTCCGGCGCGCGCTACGCGGAGAACATCTTCGGGGCTCGCGGGGCGC 289

```

36 .....TyrArgTrpArgArgLysAsnGlyIlePheAsnt 47
   |||||
290 GCCTGGAGCGCGCGACGCGTGCCTG..... 320
47 hrArgLeuSerArgThrPheGlyTyrThrValLysArgThrThrValThr 63
   |||||
321 .....GGTGGGCGGAGAGAACTACCACTAGG 350
64 ThrProSerTrpAlaValAspMetMetArgPheLysIleAspPheVa 80
   |||||
351 ACACCAACACGTGCAC.....CGGGCAAGGTGTGCGGCCACTT 391
80 lProProGlyGly.....GlyThrAsnLysIleSerIleProp 93
   |||||
392 ACACCCAGGTGTGCGCAAGTCGGTGCACATCGCTG..... 431
93 heGluTyrTyrArgIleArgLysValLysValGluPheTrpProCysSer 109
   |||||
432 .....CGCCGCGTGTGTGCG..... 449
110 ProIleThrGlnGlyAspArgGlyValGlySerThrAlaValIleLeuA 126
   |||||
450 .....GGCGAACCGCGCGCTG..... 465
126 spAspAsnPheValThrLysAlaThrAlaLeuThrTyrAspProTyrVal 142
   |||||
466 .....TTCATCACC.....TGCAACTAGGACCCCGCGGC 495
143 AsnTyrSerArgHisThrIleProGlnProPhe 154
   |||||
496 AACTTCAACGCGGAG.....CGCCGCTTC 519

```

seq_name: gb_gss10:AQ638908

```

seq_documentation_block:
LOCUS AQ638908 316 bp DNA GSS 08-JUL-1999
DEFINITION 927P1-14H7.TP 927P1 Trypanosoma brucei genomic clone 927P1-14H7,
genomic survey sequence.
ACCESSION AQ638908
NID 95115618
VERSION AQ638908.1 GI:5115618
KEYWORDS GSS.
SOURCE Trypanosoma brucei.
ORGANISM Trypanosoma brucei
Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
REFERENCE 1 (bases 1 to 316)
AUTHORS El-Sayed,N., Zhao,S., Zhao,H., Gill,S., Suh,E., Malek,J., Fujii,C.,
Gerrard,C., Leach,V., de Jong,P., Ullu,E., Melville,S.,
Donelson,J., Fraser,C. and Adams,M.
Determination of clone end sequences from Trypanosoma brucei TREU
927/4 P1 library
Other GSSs: 927P1-14H7.TV
UNPUBLISHED (1999)
CONTACT: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@tigr.org
For clone/filter availability, please contact Sara Melville
(sml60@mol.bio.cam.ac.uk). P1 end sequences search page:
http://www.tigr.org/tdb/mdb/tbdb/.
Seq primer: SP6
Class: P1 ends.

```

FEATURES
Source

Location/Qualifiers
1..316
/organism="Trypanosoma brucei"
/strain="TREU927/4"
/db_xref="taxon:5691"
/clone="927P1-14H7"
/clone_lib="927P1"

/note="Vector: pAD10SacBII; Site_1: Bam HI; Constructed by Sara Melville, University of Cambridge, UK and Nancy Shepherd, Dupont Merck, Wilmington, DE. Genomic DNA was isolated from Trypanosoma brucei (stock TREU927/4) and partially digested with Sau 3AI. DNA fragments were cloned into the Bam HI site of pAD10SacBII vector (Genbank accession U09128). The average insert size is 65 Kb. Coverage: approx 4.4 x the haploid non-minichromosomal genome"

BASE COUNT 90 a 74 c 69 g 82 t 1 others
ORIGIN

alignment_scores:
Quality: 86.00 Length: 49
Ratio: 2.867 Gaps: 1
Percent Similarity: 61.224 Percent Identity: 38.776

alignment_block:

US-09-209-961-5 x AQ638908/rev ..

Align seg 1/1 to reverse of: AQ638908 from: 1 to: 316

```

65 ProSerTrpAlaValAspMetMetArgPheLysIle..... 76
|||||
305 CCTTCGTGGCCCTTCGACGATTCCTTCGAGTCGAGCAACTCAAGA 256
77 ....AspAspPheValProGlyGlyThrAsnLysIleSerIleP 92
|||
255 TTAGGATGAGGAGTCCCGGAGGCGATGCGAAGCCGTCGACCAATCG 206
92 ropheGluTyrTyrArgIleArgLysValLysValGluPheTrpPro 107
|||||
205 AGTTCAACTCAGCGGATCATGAAAGTAGCTTCTTAGACTACCCCT 159

```

seq_name: gb_est31:AI673888

seq_documentation_block:

```

LOCUS AI673888 630 bp mRNA EST 19-MAY-1999
DEFINITION 605038f09.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays
CDNA, mRNA sequence.
ACCESSION AI673888
NID 94874368
VERSION AI673888.1 GI:4874368
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Zea.
REFERENCE 1 (bases 1 to 630)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
UNPUBLISHED (1999)
JOURNAL University
COMMENT On Mar 10, 1998 this sequence version replaced gi:2948185.

```

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 605038 row: F column: 09.

FEATURES
Source

Location/Qualifiers
1..630
/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="pericarp, embryo, and endosperm"
/dev_stage="10 days post-pollination"

/lab_host="DH5(alpha)"
 /note="Organ: Kernel; Vector: PAD-GAL4-2"; Site_1: EcoRI;
 Site_2: XhoI; kernel endosperm cDNA library from Schmidt
 lab"

BASE COUNT 131 a 185 c 239 g 74 t 1 others
 ORIGIN

alignment_scores:
 Quality: 88.50 Length: 135
 Ratio: 1.609 Gaps: 6
 Percent Similarity: 40.741 Percent Identity: 25.926

alignment_block:
 US-09-209-961-5 x AI673888/rev ..
 Align seg 1/1 to reverse of: AI673888 from: 1 to: 630

```

1 MetThrTyrProArgArgArgTyrArgArgArgHisArgProArgse 17
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
541 CTACCTTCCTCCCTCCGGCGG...AGCGGAGGAGCGGAGCGCCGTCA 495
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
17 rHisLeuGlyGln.....lLeuArgArgArgProTrp. 28
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
494 TCTACTCGTCCCGCCACCTCCTGCTCTCCGCGCGCGCGCTTGGT 445
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
29 .....LeuValHisProArgHisArgTyrArgTyr 38
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
444 GCGGAGGCTTCTACTACCGCGCCACCGCGCGGAGCGCGCAGTGG 395
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
39 ArgArgLysAsnGlyLePheAsnThrArgLeuSerArgThrPheGlyTy 55
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
394 CGG..... 392
55 rThrValLysArgThrThrValThrThrProSerTrpAlaValAspMetM 72
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
391 .....TGGCGCGTCCGACGAG 376
72 etArgPheLysIleAspPheValProProGlyGlyGlyThrAsnLy 88
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
375 GCGCGTGGTGTACCGACGCGCGCGCGCGCGCGCCCAACCGTTCCT 326
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
88 sIleSerIleProPheGluTyrTyrArgIleArgLysValValGluP 105
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
325 GCCCTACCTCCCTCTCTCTCTACAGC..... 298
105 heTrpProCysSerProIleThrGlnIleAspArgGlyValGlySerThr 121
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
297 .....CCGCGCGCGCGTGTGCACTCCGCGCGCGCGCGCGCCCTGCGCCACC 253
122 Ala 122
   :::::
252 TCG 250
```

seq_name: gb_est36:AI895135

seq_documentation_block:
 LOCUS AI895135 600 bp mRNA EST 27-JUL-1999
 DEFINITION EST264578 tomato callus, TAMU Lycopersicon esculentum cDNA clone
 CLEC6L6, mRNA sequence.

ACCESSION AI895135

NID 95601037

VERSION AI895135.1 GI:5601037

KEYWORDS EST.

SOURCE tomato.

ORGANISM Lycopersicon esculentum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
 Potatoe; Lycopersicon.

1 (bases 1 to 600)

REFERENCE Alcala,J., Vrebalov,J., White,R., Matern,A.L., Vision,T.,

Holt,I.E., Liang,F., Upton,J., Ronning,C.M., Craven,M.B.,

Fujii,C.Y., Bowman,C.L., Niernan,W., Fraser,C.M., Venter,J.C.,
 Martin,G.B., Tankley,S.D. and Giovannoni,J.
 Generation of ESTs from tomato callus tissue
 Unpublished (1999)
 On Dec 20, 1995 this sequence version replaced gi:1136036.

TITLE
 JOURNAL
 COMMENT

Contact: David Frisch
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 4366
 Fax: 864 656 4293
 Email: dfrisch@CLEMSON.EDU
 5 prime sequence.

FEATURES
 Location/Qualifiers

source

1..600
 /organism="Lycopersicon esculentum"
 /cultivar="TA496"
 /db_xref="taxon:4081"
 /map="15q26.1: 8p21.3-p22"
 /clone="CLEC6L6"
 /clone_lib="tomato callus, TAMU"
 /tissue_type="callus"
 /dev_stage="combined undifferentiated and shooting callus"
 /lab_host="XLI-Blue MRF"
 /note="Vector: pBlueScript SK(-); Site_1: EcoRI; Site_2:
 XhoI; CLEC: Tomato Callus EST Library. Directionally
 cloned cDNAs inserted into pBlueScript SK(-) at 5' end
 with EcoRI and 3' end with XhoI site."
 BASE COUNT 191 a 134 c 105 g 169 t 1 others
 ORIGIN

alignment_scores:

Quality: 87.50 Length: 223
 Ratio: 0.911 Gaps: 11
 Percent Similarity: 43.049 Percent Identity: 21.525

alignment_block:

US-09-209-961-5 x AI895135

Align seg 1/1 to: AI895135 from: 1 to: 600

```

42 AsnGlyIlePheAsnThrArgLeuSerArgThrPheGlyTyrThrVally 58
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
4 AATGGTCGCTATATATATACGTCACGCAATAAACTAATATCTTTTCC 53
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
58 sArgThrThrValThrThrPro.....SerTrpAlaValAspM 71
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
54 TAGACCTACTGCTGCAATTCCTCATATATAGGGGAGTGGTGGAAAGTG 103
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
71 etMetArgPheLysIleAspAspPheValProProGlyGlyGlyThrAsn 87
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
104 ATGTACAAGCAGTGTGCAATGAGTTTCTTACAAGTGGGAGCAGATGCAAT 153
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
88 LysIleSerIleProPheGluTyrTyrArgIleArgLysValLysValG 104
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
154 AAA.....TCTGATGCATTTCTTATAAATGCTCAACCCAGGAGA 191
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
104 uPheTrpProCysSerProIleThrGlnGlyAspArgGlyValGlySerT 121
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
192 TTTATATCCTTGCTCA..... 207
121 hrAlaValIleLeuAspAspAsnPheValThrLysAlaThrAlaLeuThr 137
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
208 .....ACAATGACACATC.....AAACTACG 231
138 TyrAspProTyrValAsnTyrSerSerArg..... 147
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
232 GTGGAGAAAAAATAGACATATTTCTTCATGCTTAACGCCGTAATGAA 281
148 .....HisThrIlePro..... 151
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
```

```

282 TAAATCTCATGTTTCTCGGTGCAATCAACCACTCACCGTCGTAGGAA 331
152 .....GlnProPheSerTyrHisSerArgTyrPheThr 162
332 CTGAGGGTGGCTACGTTAAACGTTCAAG.....TCAATATACATCACA 375
163 ProLysProValLeuAspSerThrIleAspTyrPheGlnProAsnAsnLys 179
376 ATTTCCTCCG.....GGTGAACTATTGAT..... 399
179 sArgAsnGlnLeuTrpLeuArgLeuGlnThrSerGlyAsnValAspHisV 196
400 .....GTTTACTCGAAGCAAAATCAAAACCCCTAACCACT 433
196 alGlyLeuGlyThrAlaPheGluAsnSerLysTyrAspGlnAspTyr... 211
434 ATTATATGGGCGCGCAAGCCTACATAGCGTAGTCGGAGTTATTATAC 483
212 AsnIleArgValThrMetTyrValGlnPheArg...GluPheAsnLeuLys 227
484 AACATAACCAACCGCCGCGATCGTTCAGTATCGTGGAAATTAATCACTTC 533
227 sAspProProLeuGluPro 233
534 ATCACCACCCATCCTTCCG 552
seq_name: gb_est27:AI460429

```

```

seq_documentation_block: 387 bp mRNA EST 09-MAR-1999
LOCUS sa81d12.v1 Gm-cl004 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-cl004-5712 5' similar to TR:Q39882 Q39882 NODULIN-26. ; mRNA
sequence.

```

```

ACCESSION AI460429
NID 94313310
VERSION AI460429.1 GI:4313310
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max

```

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Glycine.

1 (bases 1 to 387)

Shoemaker, R., Briner, D., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and Wilson, R.

```

TITLE Public Soybean EST Project
JOURNAL Unpublished (1999)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189388.

```

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or info@genomesystems.com web site: www.genomesystems.com
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 280.

```

FEATURES
source Location/Qualifiers
1..387
/organism="Glycine max"
/db_xref="taxon:3847"

```

```

/clone="GENOME SYSTEMS CLONE ID: Gm-cl004-5712"
/clone_lib="Gm-cl004"
/tissue_type="root"
/lab_host="XL10-Gold"
/notes="vector: pBluescript II XR; Site_1: EcoRI; Site_2: XhoI; Root cDNA. The mRNA was isolated from entire roots of 8 day old 'Williams' seedlings which were propagated on paper towels with distilled water. Stratagene's cDNA Synthesis Kit (catalog #200401) was used to synthesize the cDNA. First-strand synthesis was performed with 5-methyl dCTP, hence the ligated cDNA is hemimethylated. Stratagene's first-strand synthesis primer was used [GAGAGAGAGAGAGAGAGAGTCTCGAG(T)-18]. After second-strand synthesis, the cDNA ends were 'polished' with clone Pfu DNA polymerase, ligated to EcoRI adapters, and phosphorylated. The XhoI site within the first-strand synthesis primer was restricted by digestion with XhoI; all XhoI sites in the cDNA would be protected by their hemimethylated status. The cDNA constructs were size-fractionated with a 500bp cutoff, using GibcoBRL Life Technologies' cDNA Size Fractionation column. The column eluent was then ligated into Stratagene's pBluescript II XR Predigested vector (pBluescript II SK(+)) that had been digested with EcoRI and XhoI, and phosphorylated. Both the white and blue colonies appear to contain recombinant plasmids with cDNA inserts. Blue colonies 9n-15) have been sequenced, and possess putative cDNA inserts. This library was constructed by Dr. Paul Keim & Virginia H. Coryell, Department of Biology, Box 5640, Northern Arizona University, Flagstaff, AZ 86011, Phone: 520-523-1078 (Dr. Paul Keim), 520-523-1372 (Virginia H. Coryell), Fax: 520-523-7500, email: paul.keim@naui.edu, virginia.coryell@naui.edu"
BASE COUNT 54 a 139 c 92 g 101 t 1 others
ORIGIN

```

```

alignment_scores:
Quality: 85.00 Length: 79
Ratio: 2.024 Gaps: 2
Percent Similarity: 53.165 Percent Identity: 31.646

alignment_block:
US-09-209-961-5 x AI460429 ..
Align seg 1/1 to: AI460429 from: 1 to: 387

5 ArgArgArgTyrArgArgArgHis.....ArgProAr 16
|||||  |||||||  |||||||  |||  |||
57 CGTCGCGTCTCGTCGGCGCAACATCTCCGGCGCCACGTCACCCCG 106
|||||  |||  |||  |||  |||  |||  |||  |||  |||
16 gSerHisLeuGlyGlnIleLeuArgArgArgProTrpLeuValHisProA 33
|||||  |||  |||  |||  |||  |||  |||  |||  |||
107 CCGTCACTTCGGCGC...CTTCGTCGGCGCAACATCACCTTCCTCCGC 153

33 rghisArgTyrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeu 49
|||||  |||  |||  |||  |||  |||  |||  |||  |||
154 GGCATCGTCTACCTCATCAGCCAGCTCTCGGTCATCGTCGCTCCCT 203
|||||  |||  |||  |||  |||  |||  |||  |||  |||
50 SerArgThrPheGlyTyrThrValLysArgThrThrValThrProSe 66
|||||  |||  |||  |||  |||  |||  |||  |||  |||
204 CCTCTGGCCTTCGTACCCGCTCCACTGTTCAGCATTCGGACTCTCCG 253
|||||  |||  |||  |||  |||  |||  |||  |||  |||
66 rTrpAlaValAspMetMetArgPheLysIleAspAsp 78
|||||  |||  |||  |||  |||  |||  |||  |||  |||
254 CTGAGTGTGGAGTGGCAACGCTTGTTGTTGGAGAT 290
|||||  |||  |||  |||  |||  |||  |||  |||  |||

```

seq_name: gb_est32:AI736822

```

seq_documentation_block: 493 bp mRNA EST 14-JUN-1999
LOCUS AI736822
DEFINITION sb34a05.v1 Gm-cl012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl012-297 5' similar to TR:022525 022525 TONOPLAST INTRINSIC

```


and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening.

BASE COUNT 108 a 234 c 246 g 131 t
ORIGIN

alignment_scores:
Quality: 87.00 Length: 135
Ratio: 1.261 Gaps: 7
Percent Similarity: 51.11 Percent Identity: 28.89

alignment_block:

US-09-209-961-5 x AQ577736 ..

Align seg 1/1 to: AQ577736 from: 1 to: 719

```

6 ArgArgTyrArgArgArgArgHisArgProArgSerHisLeuGlyGlnI 22
|||||
171 CGCGGGGTGGCGCGCGAGTTCGT.....CACCTCGGCGACCT 211
|||||
22 eLeuArgArgArgProTrpLeuValHisPro..... 32
|||||
212 TCCTCATCGACGGCCACACATGGAACATCCAGGTCTACCCCGACAGGTGG 261
|||||
33 .....ArgHisArgTyr..... 36
|||||
262 AAGCAGGAGATGACGGCGGTAGTGTCTCTCTCTCTCTCTCTCTCTCGG 311
|||||
37 ArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThrPh 53
|||||
312 CGGAGCAACCGCGTAGGCGAAGTACACGCTGAGCCCTTCGGAGAACGG 361
|||||
53 eGlyTyrThrValLysArgThr.....ThrValThrThrP 65
|||||
362 CGCGAGATCCGTACACAGGAGCTTGACGATCGGTTCGACACGCGTGGCG 411
|||||
65 roSerTrpAlaValasp.....MetMetArgPheLysIleaspPhe 79
|||||
412 CTTCTCGGATTCCTCCAGGTTCTATGGAGAGGCCAGGCTCGGCAATGG 461
|||||
80 Val.....ProProGlyGlyGlyThrAsnLysIleSerIleProPh 93
|||||
462 CTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 511
|||||
93 eGlyTyrTyrArgLysValLysValLysValGlu.PheTrpProCysSer 109
|||||
512 CTCCTCAGCGTCTACAGGAGCGCTCGACCGAGGCGTGGCGCGCGCG 561
|||||
110 Pro 110
562 CCG 564

```

seq_name: gb_est37:AI946727

seq_documentation_block: 623 bp mRNA EST 17-AUG-1999
LOCUS AI946727
DEFINITION bs30a08.y1 Drosophila melanogaster adult testis library Drosophila melanogaster cDNA clone bs30a08 5', mRNA sequence.

ACCESSION AI946727
NID 5737153
VERSION AI946727.1 GI:5737153
KEYWORDS EST.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

1 (bases 1 to 623)
Andrews, J., Bouffard, G. and Oliver, B.
Drosophila melanogaster testis expressed sequence tags
Unpublished (1999)
On Jun 5, 1998 this sequence version replaced gi:3189407.

Contact: Brian Oliver
Laboratory of Cellular and Developmental Biology
NIDDK, National Institutes of Health
6 Center Drive MSC 2715, Bldg 6, Rm B1-13, Bethesda, MD 20892 USA
Fax: (301) 496 5239
Email: oliverbhelix.nih.gov,
http://www.niddk.nih.gov/intram/people/boliver.htm

Tissue isolation and library construction performed at the National Institute of Diabetes and Digestive and Kidney Diseases, NIH (see http://www.niddk.nih.gov/intram/people/boliver.htm). DNA sequencing and analyses performed by National Institutes of Health Intramural Sequencing Center (NISC; see http://www.nisc.nih.gov). Chromatogram data were analyzed and evaluated for high quality using the ted program (Gleeson T and Hillier L, 1991).
Plate: 30 row: a column: 08
Seq primer: M13Rpl reverse primer (ABI).

FEATURES

source

```

1..623
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y[*] w[67cl]/Y"
/db_xref="taxon:7227"
/map="703C03; 1"
/clone="bs30a08"
/clone_lib="Drosophila melanogaster adult testis library"
/sex="male"
/dev_stage="1-5 day adult"
/lab_host="SOLR (Stratagene)"
/note="Organ: testis; Vector: pBlueScript SK (Stratagene); Site:1: EcoR I; Site:2: Xho I; Testes dissected from 1-5 day adult y[*] w[67cl]/Y males raised at 25°C. RNA isolated using Trizol (Life Technologies) and a single round of Poly(A)+ selection using Oligotex (Qiagen). cDNA library constructed using Stratagene ZAP-cDNA synthesis kit. Oligo dt-primed, size fractionated -1-6 kb, and directionally cloned at EcoRI and XhoI in Uni-ZAP XR. Following a single round of amplification pBlueScript SK phagemids were mass excised. A distribution channel for clones is being sought, but not currently available. Requests for clones cannot be honored."
BASE COUNT 189 a 141 c 184 g 109 t
ORIGIN

```

alignment_scores:
Quality: 86.00 Length: 135
Ratio: 1.211 Gaps: 6
Percent Similarity: 52.593 Percent Identity: 29.630

alignment_block:
US-09-209-961-5 x AI946727 ..

Align seg 1/1 to: AI946727 from: 1 to: 623

```

4 ProArgArgTyrArgArgArgHisArgProArgSer...HisLe 19
|||||
259 CGCGGAATCGCCCAAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 308
|||||
19 uGlyGlnIleLeuArgArgArgProTrpLeuValHisProArgHisArg 36
|||||
309 AAGCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 346
|||||
36 yrArgTrpArgArgLysAsnGlyIlePheAsnThrArgLeuSerArgThr 52
|||||
347 GGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 393
|||||
53 PheGlyTyrThrValLysArgThrThrValThrProSerTrpAlaVa 69
|||||

```

```

394 .....AGAAAGAAAAGAGGAGAGAGAGGAGCCCAAGGGCAA 428
69 1aspMetMetArgPheLysIleAspPheValProProGlyGlyT 86
   ||| ..... |||||..... |||||..... |||||.....
429 GAGAAGGCAAGGAGGCAAGTACGACATATAGACCCCAACGACATG 478
   ||| ..... |||||..... |||||..... |||||.....
86 hrAsnLysIleSerIle...ProPheGluTyrTyrArgIleArgLysVal 101
   ||| ..... |||||..... |||||..... |||||.....
479 TCAAGGAGCAGGATGTTTCCCAAAATTCATCGAATCCATCGTT 528
   ||| ..... |||||..... |||||..... |||||.....
102 .....LysValGluPheTrpProCysSerProIleTh 112
   ||| ..... |||||..... |||||..... |||||.....
529 TCTAACTTCAAAATTTCAAGTGCATATATATGCGATACCTAACGATAAC 578
   ||| ..... |||||..... |||||..... |||||.....
112 rGlnGlyAspArgGlyValGlySerThrAlaValIleLeuAspAspAsp 129
   ||| ..... |||||..... |||||..... |||||.....
579 TATTGGTCAG.....TCCATTGCCCGTTGATGGAAGGCTTCT 616
   ||| ..... |||||..... |||||..... |||||.....
129 heVal 130
   ||| ..... |||||..... |||||..... |||||.....
617 TTATT 621

seq_name: gb_est36:AI900445
seq_documentation_block: 337 bp mRNA EST 27-JUL-1999
LOCUS AI900445
DEFINITION SC05h05.Y1 Gm-cl012 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl012-1330 5' similar to TR:Q41616 Q41616 GAMMA-TIP-LIKE PROTEIN
; mRNA sequence.
ACCESSION AI900445
NID G5606347
VERSION AI900445.1 GI:5606347
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae;
Glycine.
1 (bases 1 to 337)
Shoemaker, R., Briner, D., Mairra, M., Hillier, L., Kucaba, T.,
Martin, J., Beck, C., Wylie, T., Underwood, K., Steptoe, M.,
Theising, B., Allen, M., Bowers, Y., Person, B., Swaller, T.,
Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Wilson, R.
Public Soybean EST Project
Unpublished (1999)
On Jun 22, 1998 this sequence version replaced gi:3246831.

Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
Trace considered overall poor quality
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 1.
Location/Qualifiers
1. .337
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl012-1330"
/tissue_type="Apical shoot tips, 9-10 day old etiolated

394 .....AGAAAGAAAAGAGGAGAGAGGAGGAGCCCAAGGGCAA 428
/lab_host="XL10-Gold"
/Note="Vector: pBluescript II XR; Site_1: EcoRI; Site_2:
XhoI; This cDNA library was constructed from mRNA isolated
from the apical shoots of 9 to 10 day old etiolated
seedlings. The shoot tips including any emerged leaves
were harvested for mRNA isolation. The cDNA library was
prepared using the Stratagene pBluescript II XR cDNA
library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments were
transformed into XL10-Gold host cells."
BASE COUNT 48 a 140 c 74 g 74 t 1 others
ORIGIN

alignment_scores:
Quality: 82.50 Length: 74
Ratio: 2.230 Gaps: 5
Percent Similarity: 50.000 Percent Identity: 39.189

alignment_block:
US-09-209-961-5 x AI900445 ..
Align seg 1/1 to: AI900445 from: 1 to: 337
5 AtgArgArgTyrArgArgArgHis.....ArgProAr 16
||||| |||||..... |||||
124 CGTCGGCGTCTCGTCGGCGCAACATCTCCGGGTCAGTCAACCCG 173
16 gserHisLeuGlyGlnIleLeuArgArgProTrpLeuValHis.... 31
||| ||||| |||||
174 CGGTCA.....CTTCGGCGCTTCGTGGCGCAACATCACC 211
32 .....ProArgHisArgTyrArgTrpArgLysAsnGlyIlePheAsn 46
||||| ||||| |||||
212 TTCCTCCGGCGCATCTCTACGTATCGCCAGCT..... 246
47 ThrArgLeuSerArg_ThrPheGlyTyrThrValLysArgThrValT 63
||||| ||| :||| |||||
247 CCTCGGCTCATCGTACGTCCTCTCTCTGCTCCGACCGACTGCA 296
63 hrThrProSerTrpAlaVal 69
|||:|||||:.....
297 CTGTGGCAGCATTCGGACTC 316

seq_name: gb_gss2:CNS017HI
seq_documentation_block:
LOCUS CNS017HI 915 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC
BACN17L03 of DrosBAC library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL108048
NID G5628352
VERSION AL108048.1 GI:5628352
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 915)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a

```


371 GATGAGGACGAGGGCGTGGCAGGCGGGGTCTGGGCGGTTCGTCCA 322

Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 371.

FEATURES

Source

1. .394
Location/Qualifiers
/organism="Glycine max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl017-761"
/clone_lib="Gm-cl017"
/tissue_type="vegetable buds of field grown plants"
/lab_host="XL10-Gold"
/notes="Vector: pBluescript II XR; Site 1: EcoRI; Site 2:
XhoI; This cDNA library was constructed from mRNA isolated
from vegetable buds of field grown plants. The cDNA
library was prepared using the Stratagene pBluescript II
XR library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a poly
(dT) sequence with a XhoI restriction site. EcoRI
adapters were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA fragments were
directionally cloned into the EcoRI-XhoI restriction site
of the pBluescript vector. The ligated cDNA fragments
were transformed into XL10-Gold host cells."

BASE COUNT 66 a 156 c 89 g 83 t
ORIGIN

alignment_scores:
Quality: 82.00 Length: 40
Ratio: 3.280 Gaps: 2
Percent Similarity: 62.500 Percent Identity: 52.500

alignment_block:

US-09-209-961-5 x AI941495 ..

Align seg 1/1 to: AI941495 from: 1 to: 394

5 ArgArgGTYrArgArgHis.....ArgProAr 16
||||||| |||||||:|||||
258 CGTCGCGTCTCCGTCGCGCCACATCTCCGCGCGCCACGTCACCCCG 307
|||
16 gSerHisLeuGlyGlnIleLeuArgArgProTrpLeuValHisProA 33
| ||||| :|||
308 CCGTCACGTCGCGGC...CTTCGTCGCGCCACATCACCTCTCTCCGC 354
|||
33 rgHisArgTYrArgTrpArg 39
||||||| ||| |||
355 GGCATCGCTACGTCATCGC 374


```

/db_xref="GI:3668365"
/translation="MVLNCCPSCRSSLSSEDFLGLILEECYKTIHGGRGVRHPPPPM
P"
CDS
1016..1177
/codon_start=1
/product="P6.2"
/protein_id="AAC61863.1"
/db_xref="PID:93668366"
/db_xref="GI:3668366"
/translation="MVFIIHLGFKWGVKIKFSELYIRGYTDIVLVVFTVFERSAEA
YVHISRGL"
CDS
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61864.1"
/db_xref="PID:93668367"
/db_xref="GI:3668367"
/translation="MTYPRRRRRRRRSHLQILRRRRLVHPHRYRWRKNGI
FNTLSRTFGTVKRTVTTPSWAVDMRFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTIDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLDSITDYFQPNKRNQLRLQTSNVHDVHGLGTAFENSKYDQDYNIRVTMY
VQFRENLKDPLEP"
BASE COUNT 463 a 362 c 481 g 462 t
ORIGIN

```

```

alignment_scores:
  Quality: 311.00      Length: 59
  Ratio: 5.271        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-209-961-7 x AF085695/rev ..

Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
565 ATGAAGTGACATTGGTTTCCCAATCAGCGTTTTCGATTTCCCGCTCAC 516

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
515 TTTCAAAGTTCAGCCAGCGCGGAAATTTTGACAAAGTTACAGGGT 466

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
465 GCTGCTTTGCAAGCGTCACGAAATCCCGCTCTCCAAAGGTTACTCACA 416

51 AlavAlaSpArgSerLeuArgCysPro 59
|||||
415 GCAGTAGACAGGTCACCTCGTTGTCCT 389

```

seq_name: gb_v1:AF086834

```

seq_documentation_block:
LOCUS AF086834 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain B9, complete genome.
ACCESSION AF086834

```

```

NID 93661515
VERSION AF086834.1 GI:3661515

```

```

KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus.

```

```

REFERENCE
1 (bases 1 to 1768)
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.

```

```

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Emergence of a new porcine circovirus
JOURNAL Unpublished

```

```

REFERENCE
2 (bases 1 to 1768)

```

```

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Direct Submission
JOURNAL Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

```

```

FEATURES
source
1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
51..995

```

CDS

```

/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:93661516"
/db_xref="GI:3661516"
/translation="MPSKNGRSQPQPHRWVFTLNNPSEDERKKIRELPISLFYFI
VGEENEGRTPHLQGFANFKKVFKNKYLKGARHIEKAGKTQQQNKCYCSKEGN
LLIEGAPRSQSDLSSTAVTLLSEGLITVAKQHPVTFKNFRGLAEKLVSGKM
OKRDKTNVHFIVPGPGCGKSKAANFANPETTYKPKKNKWDGYHGEKVVYDDFY
GWLPHDDLRLCDRYPVLTAKGTGTPFLARSILITSNOTPLEWYSSSTAVPAVEALYR
RITSLVFNWNTQEQTSEGGQFVTLSPCPPEPPEYIN"
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:93661517"
/db_xref="GI:3661517"
/translation="MTYPRRRRRRRRSHLQILRRRRLVHPHRYRWRKNGI
FNTLSRTFGTVKRTVTTPSWAVDMRFKIDDFVPGGNTKISIPFEYIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVKATALTIDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLDSITDYFQPNKRNQLRLQTSNVHDVHGLGTAFENSKYDQDYNIRVTMY
VQFRENLKDPLEP"

```

CDS

```

BASE COUNT 463 a 357 c 482 g 466 t
ORIGIN

```

alignment_scores:

```

  Quality: 311.00      Length: 59
  Ratio: 5.271        Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

```

alignment_block:

US-09-209-961-7 x AF086834/rev ..

Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
565 ATGAAGTGACATTGGTTTCCCAATCAGCGTTTTCGATTTCCCGCTCAC 516

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
515 TTTCAAAGTTCAGCCAGCGCGGAAATTTTGACAAAGTTACAGGGT 466

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
465 GCTGCTTTGCAAGCGTCACGAAATCCCGCTCTCCAAAGGTTACTCACA 416

51 AlavAlaSpArgSerLeuArgCysPro 59
|||||
415 GCAGTAGACAGGTCACCTCGTTGTCCT 389

```

seq_name: gb_v1:AF086835

seq_documentation_block:

```

LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION AF086835

```

```

NID 93661518
VERSION AF086835.1 GI:3661518

```

```

KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus.

```

```

REFERENCE
1 (bases 1 to 1768)
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.

```

```

AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
TITLE Emergence of a new porcine circovirus

```



```

LOCUS       AF118095             597 bp            DNA                VRL                27-JUL-1999
DEFINITION  Porcine circovirus type 2 isolate IAF-614 unknown gene.
ACCESSION   AF118095
NID         95596425
VERSION     AF118095.1  GI:5596425
SOURCE      porcine circovirus type 2.
ORGANISM    porcine circovirus type 2
VIRUSES:    ssDNA viruses; Circoviridae; Circovirus.
REFERENCE   1 (bases 1 to 597)
AUTHORS     Ouardani,M., Wilson,L., Jette,R., Jette,R., Montpetit,C. and Dea,S.
TITLE       Multiplex PCR for detection and typing of porcine circoviruses
JOURNAL     J. Clin. Microbiol. (1999) In press
REFERENCE   2 (bases 1 to 597)
AUTHORS     Ouardani,M., Wilson,L., Jette,R., Jette,R., Montpetit,C. and Dea,S.
TITLE       Direct Submission
JOURNAL     Submitted (06-JAN-1999) CRMb, INRS-IAF, 531, Boulevard des
            Prairies, Laval, Que H7N 4Z3, Canada
FEATURES             Location/Qualifiers
     source          1..597
                     /organism="porcine circovirus type 2"
                     /isolate="IAF-614"
                     /db_xref="taxon:85708"
     note            "virus isolated from lung, lymph node, spleen and
                     tonsil tissue from pigs affected by postweaning
                     multisystemic wasting syndrome"
     CDS             <1..>597
                     /note="ORF1"
                     /codon_start=1
                     /product="unknown"
                     /protein_id="AAD45580.1"
                     /db_xref="PID:95596426"
                     /db_xref="GI:5596426"
                     /translation="SEDERKKIRLEPISLFDYFVGEENEGRTPHLQGFANFVKQ
                     TENKVKYLGIAARHLEKAGTDHONKCYSGENLLIECGAPPSQGRSDLSPAVSTL
                     LKESGLTVVAEHPPTFVFNFRGLAELLKVSQKMRDKNTVHVIVGPPGCGKSKWA
                     ANFADPTTIVAPPRKNKWDVGHGEVVVIDDFYGLWLPW"
     BASE COUNT      169 a   112 c   170 g   146 t
     ORIGIN
alignment_scores:
    Quality: 295.00      Length: 59
    Ratio: 5.175        Gaps: 0
    Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF118095/rev ..

Align seg 1/1 to reverse of: AF118095 from: 1 to: 597

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
443 ATGACGGTGATCATTTGGCTTCCCAATCAGCTTCTGCATTTCCCGCTCAC 394

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
393 TTTCAAAGTTCAGCGACCGCCGCGGAATTTCTGACAAACGTTACAGGGT 344

34 ysCysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
343 GCTGCTCTGC AACGGTCA CACAGACTCCGCGCTCTCCAAACAGGTACTCACA 294

51 AlavalaspargSerLeuArgCysPro 59
|||||
293 GCAGTAGACAGGTCACTCCGTTGCCT 267

seq_name: gb_vi:AF055393
seq_documentation_block:
LOCUS       AF055393             1767 bp            DNA                circular VRL                13-SEP-1998
DEFINITION  Porcine circovirus Type II from France, complete genome.
ACCESSION   AF055393

```

```

/protin_id="AAC35322.1"
/db_xref="PID:g3598823"
/db_xref="GI:3598823"
/translation="MTIPILVSRWPCGFRVCKISSPAFTTPRPHNDVYLSLPI
TLHFAHFAKFSQPAISDKRYRLVLCNGHQTPALQGGHSSRQVTPLSLRSSSL
HQ"
complement(733..912)
/notes="predicted 6.5 kda protein"
/codon_start=1
/product="ORF4"
/protin_id="AAC35323.1"
/db_xref="PID:g3598824"
/db_xref="GI:3598824"
/translation="MTCLVFSQRFICPLTFKSSASPRFLTNVTCSSATVRLPL
SNKVLAVDRSLRCP"
complement(1035..1100)
/notes="predicted 2.3 kda protein"
/codon_start=1
/product="ORF8"
/protin_id="AAC35324.1"
/db_xref="PID:g3598825"
/db_xref="GI:3598825"
/translation="MDIDHTVSDHPRAASHKSHQ"
1330..1335
complement(1336..1380)
/notes="predicted 1.8 kda protein"
/codon_start=1
/product="ORF11"
/protin_id="AAC35326.1"
/db_xref="PID:g3598827"
/db_xref="GI:3598827"
/translation="MNNKHVEYIKKTQ"
complement(1369..1374)
complement(join(1380..1767,1..314))
/notes="predicted 27.8 kda protein"
/codon_start=1
/product="ORF2"
/protin_id="AAC35321.1"
/db_xref="PID:g3598822"
/db_xref="GI:3598822"
/translation="MTYPRRYRRRRRSHLQILRRPWLVPVRRHYWRNRKNGI
FNRLSRTFYTVKTRVTPSWAVDMRFINDELPGGGNSPRSPFVYRIRKVK
VEWPCSPITQGRGVSSAVIDNFVKTATLTDPYVNYSSRHITOPFSYHSRY
FTPKPVLDSIDIFQPNKRNLWLRLQTAGNDVHVLGTAFENSIYDQENIRVTHY
VQFRENKDPPLNP"
BASE COUNT 447 a 360 c 502 g 458 t
ORIGIN

alignment_scores:
  Quality: 295.00      Length: 59
  Ratio: 5.175        Gaps: 0
  Percent Similarity: 96.610  Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF055393/rev ..

Align seg 1/1 to reverse of: AF055393 from: 1 to: 1767

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProteinTh 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
912 ATGACGTGTACATTAGTCTTCAATCAGCGTCTGCAATTTCCCGCTCAC 863
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
862 TTTCAAAGTTCAGCCGCCCGGAAATTTCTGACAAACGTTACAGGT 813
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
812 GCTGCTCTCAACGGTCAACGACTCCCGCTCTCAACAAAGGTACTCACA 763
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 AlaValAspArgSerLeuArgCysPro 59
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```

762 GCAGTAGACAGGTCACTCCGTTGTCC 736
seq_name: gb_vi:AF055394

seq_documentation_block:
LOCUS AF055394 1767 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from France, complete genome.
ACCESSION AF055394
NID g3598831
VERSION AF055394.1 GI:3598831
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
1 (bases 1 to 1767)
MECHANISM: "porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1011 48285 EcoRI No.12"
/country="France"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protin_id="AAC35339.1"
/db_xref="PID:g3598841"
/db_xref="GI:3598841"
/translation="MASSTPASPPSDILSSPQSKRPDPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protin_id="AAC35337.1"
/db_xref="PID:g3598839"
/db_xref="GI:3598839"
/translation="MSTAQEGVLTVVRLTVTPKVRERRVLMKPFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protin_id="AAC35338.1"
/db_xref="PID:g3598840"
/db_xref="GI:3598840"
/translation="MAAGAVSSSPVPPWIRHI"
complement(311..439)
/notes="predicted 4.6 kda protein"
/codon_start=1
/product="ORF9"
/protin_id="AAC35335.1"
/db_xref="PID:g3598837"
/db_xref="GI:3598837"
/translation="MGLGSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRFQI"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protin_id="AAC35330.1"
/db_xref="PID:g3598832"
/db_xref="GI:3598832"
/translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRDLPISLFDYFI
VGEENEGERTPHLQGFANFKVKTQFNKVKYLGARCHLEKAKAGDQONKEVCSKRG
LMECGAPRSQGSDDLSTAVSTLLSESLVTVAEQHPVTVFVNRFRGLAEKLVSGRM
```

QKRDWKTNVHVIIVGPPGCKSKWAANFADPTTYWKPNNKWDGYHGEVYVDDFY
GWLDDLLRLCDRLPLTVETKGGTVFPLARSILLITNOTPLEWYSTTAVPAVALYR
RTLSLVFNKNATEQSTEEGGQFVTLSPPCPEFFEYNY

polyA_signal

CDS

complement(704..1018)
/note="predicted 11.9 kda protein"

/codon_start=1

/product="ORF3"

/protein_id="AAC35332.1"

/db_xref="PID:g3598834"

/db_xref="GI:3598834"

/translation="MYTIPPLVSRMFPVCGFRVCKISSPFAFTTPRPHNDVYISLPI
TLHFPAHFQKESQPAEISDKRYRVLNCGHOTPALQGHSSRQVTPLSLRSSTL
H"

CDS

complement(733..912)

/note="predicted 6.5 kda protein"

/codon_start=1

/product="ORF4"

/protein_id="AAC35333.1"

/db_xref="PID:g3598835"

/db_xref="GI:3598835"

/translation="MTCLVLFQSRFCIFPLTFKSSASPRKFLTNVTGCSATVTRLPL
SNKLTAVDRSLRCP"

CDS

complement(1035..1100)

/note="predicted 2.3 kda protein"

/codon_start=1

/product="ORF8"

/protein_id="AAC35334.1"

/db_xref="PID:g3598836"

/db_xref="GI:3598836"

/translation="MDIDTQSVSDHPRAASHKSHQ"

1330..1335

complement(1336..1380)

/note="predicted 1.8 kda protein"

/codon_start=1

/product="ORF11"

/protein_id="AAC35336.1"

/db_xref="PID:g3598838"

/db_xref="GI:3598838"

/translation="MNNKHVEYIKTKQ"

complement(1369..1374)

/note="predicted 27.8 kda protein"

/codon_start=1

/product="ORF2"

/protein_id="AAC35331.1"

/db_xref="PID:g3598833"

/db_xref="GI:3598833"

/translation="MTYPRRYRRRRPRSHLGLQLRRPWLHVPHRYRWRKNGI
FNTRLSRTGYTKRTVKTTPSWADMMREINDELPPGGSGNPRSPVPEYVIRKVK
VFEPGSPITQDGRGVSSAVIDDNEFVNKATLYDPIVYNSRRITQPPFSYHSRY
FTPKPVLDTIDYFQPNKRNOLWRLQTTGNVDHVLGTAFENSIIQDEYINRVTMY
VQFREPLNDPLNP"

BASE COUNT 448 a 359 c 500 g 460 t
ORIGIN

alignment_scores:

Quality: 295.00 Length: 59

Ratio: 5.175 Gaps: 0

Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:

US-09-209-961-7 x AF055394/rev ..

Align seg 1/1 to reverse of: AF055394 from: 1 to: 1767

1 MetIysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17

|||||

912 ATGAGGTGTACATTAGCTTCCAAATCAGCTTCTGCATTTTCCCGCTCAC 863

|||||

17 rPheIysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34

|||||

862 TTCTAAAGTTTCAGCAGCCGCGGAAATTTCTGACAAACGTTACAGGCT 813

34 yscysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50

|||||

812 GCTGCTCTGCAACGGTCACAGACTCCGCGTCTCCACAAGGTACTCACA 763

51 AlaValaspArgSerLeuArgCysPro 59

|||||

762 GCAGTAGACAGGTCACTCCGTTGTCCC 736

seq_name: gb_v1:AF027217

seq_documentation_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998

DEFINITION Porcine circovirus strain pmws PCV, complete genome.

ACCESSION AF027217

NID 92689645

VERSION AF027217.1 GI:2689645

KEYWORDS

SOURCE

porcine circovirus

VIRUSES; SSNA viruses; Circoviridae; Circovirus.

REFERENCE

1 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.

TITLE Nucleotide sequence of porcine circovirus associated with

postweaning multisystemic wasting syndrome in pigs

J Virol. 72 (6), 5262-5267 (1998)

JOURNAL 98241772

MEDLINE

REFERENCE

2 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.S.

TITLE Direct Submission

JOURNAL Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services

Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg,

Manitoba R3T 5S6, Canada

FEATURES

source

1..1768

/organism="porcine circovirus"

/strain="pmws PCV"

/db_xref="taxon:46221"

/note="both strands of seven overlapping PCR fragments
were sequenced; virus isolated from lung, lymph node,
spleen and tonsil tissue from pigs affected by post
weaning multisystemic wasting syndrome"

51..995

/note="ORF1; similar to Rep protein encoded by

non-pathogenic PCV, GenBank Accession Number U49186;

predicted 35.8 kda protein"

/codon_start=1

/product="putative Rep protein"

/protein_id="AAC59462.1"

/db_xref="PID:g2689646"

/db_xref="GI:2689646"

/translation="MPSKNGRSGQPQKRWVFLNPNSEDEKKIRELPISLFDYFI
VGEENEGRTPHLQGFANFKVQKOTFNKVKWLGARCYIEKAKGTDQONKEYCSKGN
LLIEGAPRSOGORSLSLSTAVSLTGESLVTVAEQHPVTFVRNFRGLAELLKYSKGM

QKRDWKTNVHVIIVGPPGCKSKWAANFADPTTYWKPNNKWDGYHGEVYVDDFY
GWLDDLLRLCDRLPLTVETKGGTVFPLARSILLITNOTPLEWYSTTAVPAVALYR

RTLSLVFNKNATEQSTEEGGQFVTLSPPCPEFFEYNY

complement(386..565)
/note="ORF4; predicted 6.5 kda protein"

/codon_start=1

misc_feature

327..332

/note="glycosylation site"

complement(357..671)

/note="ORF3; predicted 11.9 kda protein"

/codon_start=1

/protein_id="AAC59464.1"

/db_xref="PID:g2689648"

/db_xref="GI:2689648"

/translation="MTYIPPLVSRMFPVCGFRVCKISSPFAFTTPRPHNDVYISLPI
TLHFPAHFQKESQPAEISDKRYRVLNCGHOTPALQGHSSRQVTPLSLRSSTL
H"

CDS

```

/misc_feature
/db_xref="PID:g2689649"
/db_xref="GI:2689649"
/translation="WTCILVFSRRCIFPLTKFSSASPKELTNVTGCSATVTRLPL
SNKVLTAVDRSLRCP"
complement(470..478)
/feature="glycosylation site"
complement(688..753)
/feature="ORF8: predicted 2.3 kda protein"
/codon_start=1
/protein_id="AAC59469.1"
/db_xref="PID:g2689653"
/db_xref="GI:2689653"
/translation="MDIDHTVSDVDPHTAASHKSHQ"
816..824
/feature="glycosylation site"
906..914
/feature="glycosylation site"
983..988
complement(989..1033)
/feature="ORF11: predicted 1.8 kda protein"
/codon_start=1
/protein_id="AAC59472.1"
/db_xref="PID:g2689656"
/db_xref="GI:2689656"
/translation="MKNKNHYEVIKKTQ"
1016..1177
/feature="ORF5: predicted 6.2 kda protein"
/codon_start=1
/protein_id="AAC59466.1"
/db_xref="PID:g2689650"
/db_xref="GI:2689650"
/translation="WVFPHLGLKMGVFKIFSELYIHGYTDIVLVVTVFERSAEA
YVHHSRGL"
complement(1022..1027)
complement(1034..1735)
/feature="ORF2: predicted 27.8 kda protein"
/codon_start=1
/protein_id="AAC59463.1"
/db_xref="PID:g2689647"
/db_xref="GI:2689647"
/translation="MTYPRRRYRRRRRPSHLGQLRRRPWLVPHPHYRWRKNGI
FNTRLSRFGVTKATVTPSMVDMRENIDDFVPGGNTKISIPPEYVIRKVK
VEFWCSPTQDGRGVSTAVILDNFVKATALYDYPVYSSRHTIQPQSYHSRY
FYPKPLDSTIDYFPNNKRTQLMLKQTSRNVHDVGLGTAFENSIYDQDINIRVTM
VOFRFNLDKPLKP"
complement(1301..1309)
/feature="glycosylation site"
complement(1522..1611)
/feature="ORF6: predicted 3.1 kda protein"
/codon_start=1
/protein_id="AAC59467.1"
/db_xref="PID:g2689651"
/db_xref="GI:2689651"
/translation="MASSTPASAPSDILSRLPQSRPPGRWT"
1524..1631
/feature="ORF10: predicted 4.1 kda protein"
/codon_start=1
/protein_id="AAC59471.1"
/db_xref="PID:g2689655"
/db_xref="GI:2689655"
/translation="MSTAQEGVLTVALVYPKVRRRVLKMPFFLLQR"
1682..1741
/feature="ORF7: predicted 1.9 kda protein"
/codon_start=1
/protein_id="AAC59468.1"
/db_xref="PID:g2689652"
/db_xref="GI:2689652"
/translation="MAAGVSSSAVTPPWIRHS"
complement(join(1732..1768,1..92))
/feature="ORF9: predicted 4.6 kda protein"
/codon_start=1
/protein_id="AAC59470.1"

```

```

/db_xref="PID:g2689654"
/db_xref="GI:2689654"
/translation="MWLGSASSILLAGHVAAEVLPRCCRCRSALVILTAHFFRQL"
stem_loop
join(1746..1768,1..13)
/function="putative replication site"
rep_origin
join(1762..1768,1..2)
/feature="AAGTATTAC" is similar to the nonanucleotide motif
in the non-pathogenic PCV, GenBank Accession Number
U49186"
BASE COUNT 452 a 360 c 495 g 461 t
ORIGIN
alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915
alignment_block:
US-09-209-961-7 x AF027217/rev ..
Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
565 ATGACGTGTACATTGGTCTTCCAATCACGCTTCTGCAATTTCCCGCTCAC 516
17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
515 TTTCAAAAGTTACGCCAGCCCGCGGAAATTTCTGACAAACGTTACAGGTT 466
34 yScyPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
465 GCTGCTCTGCACAGGTACACGACTCCCGCTCTCCACAAGGTACTCACA 416
51 AlaValAspArgSerLeuArgCysPro 59
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
415 GCAGTAGACAGGTCACTCCGTGTGCTT 389
seq_name: gb_v1:AF055391
seq_documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from USA, complete genome.
ACCESSION AF055391
NID 93598796
VERSION AF055391.1 GI:3598796
KEYWORDS porcine circovirus.
SOURCE porcine circovirus.
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.
FEATURES
source
1..1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp. 999 EcoRI No.8"
/country="USA"
complement(101..190)
CDS

```


/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:g3598807"
/db_xref="GI:3598807"
/translation="MASSTPASPAPSDILSLPQSERPPGRWT"
103..210
CDS

/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:g3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVVLTYPKVRERRVLKMPFFLLOR"
261..320
CDS

/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"
/db_xref="PID:g3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSAVTPPWIRHS"
complement(311..439)
/note="predicted 4.6 kDa protein"
CDS

/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLSASSILLAGHVAEVLPCRCRSALVILTAHFFRFQL"
398..1342
CDS

/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKNGRSGQPQHKRWVFTLNPNSEDERKKIRELPISLFYFI
VGEENEGRTPHLOGFAFNKVKQTFNKVYLGFARCHIEKAKGTDOQNKYCSKEGN
LLIEGAPRSQSGDLSTAVSTLESGLTVAVQHPVTVFVRNFGLAELIKVSGKM
QRDKWTVHVIVGPPGGCKSKWAANFADPTTYWKPNNKWDGYHGEVVVDDFY
GWLWDLLRLCDRYPLAVETKGGTVPLARSILITSNOTPLEWTSSTPAVPAVALYR
RITSLVFWKNATEQSTEGGQVTLSPCPPEFYINY"
674..679
polyA_signal

/complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="MVTIPPLVSRWFPVCGFRVKISSPFAFTTPRWPHNDVYIGLPI
TLHHPAHQFQSPAEISDKRYVLLCNGHQTALQOQTHSSROVTPLSLRSSTF
NK"
CDS

/complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCVLVFSQRFICIFPLTFKSSAPRKFLTNVTGCCSATVTRLPL
SNKVLTVADRSURCP"
/complement(1035..1100)
/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDHPTAAASHKSHQ"
1330..1335
polyA_signal

/complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
CDS

/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKKTQ"
1363..1524
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVFKFSELYIHGYTDIVLVVVTVFERSAEA
YVYHISRGL"
polyA_site

/complement(1369..1374)
/complement(join(1381..1768,1..314))
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35299.1"
/db_xref="PID:g3598798"
/db_xref="GI:3598798"
/translation="MTYPRRYRRRRPRSHLQILRRPWLHVHPRHYRWRKNCI
FNRLSRFTGYTKATITVTPSWAVDMRFNIDDFVPPGGTNKISIPFIYRIKKVK
VEFWPCSPITQDGRGVGSTAVILDDNFVKATALYDPYVNYSSRHTIPQPFYSHSRY
FTPKPVLSTIDYFQPNKRNQLWLRLQTSRNVHVLGLTAFENSIYDQDYNIRVTMY
VQFRFNLDKDPPLKP"
BASE COUNT 453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:

US-09-209-961-7 x AF055391/rev ..

Align seg 1/1 to reverse of: AF055391 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
912 ATGACGTGTACATGGTCTTCCAATCAGCTTCGATTTCCCGCTCAC 863
17 rPheIysSerSerAlaSerProArgLysPheIeuThrAsnValThrGlyC 34
862 TTTCAAAAGTTTCAGCCAGCCCGCGGAAATTTTGACAAACGTTACAGGT 813
34 yScyPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
812 GCTGCTGTGCAACGGTTCACAGACTCCCGCTCTCCACAAGTACTCACA 763
51 AlaValAspArgSerLeuArgCysPro 59
762 GCAGTAGACAGGTCACTCCGTGTCTCT 736

seq_name: gb_vi:AF055392

seq_documentation_block:
LOCUS AF055392 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from Canada, complete genome.
ACCESSION AF055392
NID g3598808
VERSION AF055392.1 GI:3598808
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus
REFERENCES
1 (bases 1 to 1768)
Meehan,B.M., McNeilly,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
Characterization of novel circovirus DNAs associated with wasting
syndromes in pigs

```
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan.B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.

FEATURES
    Location/Qualifiers
        1..1768
            /organism="porcine circovirus"
            /specific_host="Sus scrofa"
            /strain="Porcine circovirus Type II"
            /db_xref="taxon:46221"
            /clone="Imp.1010-Stoon EcoRI No.14"
            /country="Canada"
            complement(101..190)
            /codon_start=1
            /product="ORF6"
            /protein_id="AAC35319.1"
            /db_xref="PID:g3598819"
            /db_xref="GI:3598819"
            /translation="MASTPASPAPSDILSSVPSQRPGRWT"
            103..210
            /codon_start=1
            /product="ORF10"
            /protein_id="AAC35317.1"
            /db_xref="PID:g3598817"
            /db_xref="GI:3598817"
            /translation="NSTAQGGVTVVRLTVYPKVRRRLKMPFFLLQR"
            261..320
            /codon_start=1
            /product="ORF7"
            /protein_id="AAC35318.1"
            /db_xref="PID:g3598818"
            /db_xref="GI:3598818"
            /translation="MAAGAVSSSVTPPPWIRHS"
            complement(311..439)
            /note="predicted 4.6 kda protein"
            /codon_start=1
            /product="ORF9"
            /protein_id="AAC35315.1"
            /db_xref="PID:g3598815"
            /db_xref="GI:3598815"
            /translation="MWLGSASSILLAGHVAEVLPRCCRSALVILTAHFREFQL"
            398..1342
            /codon_start=1
            /product="putative Rep protein (ORF1)"
            /protein_id="AAC35309.1"
            /db_xref="PID:g3598809"
            /db_xref="GI:3598809"
            /translation="MPSKKNRSGQPQPKRWVFTLNNPSEDERKKIRELPISLFDYFI
            VGEENEGRTPHLOGFANFVKQTFNKVYLGARCHEKAKGTDOONKEVCSKEGN
            LLIEGAPRSQSGSDLSFVSLLESGLTVVAQHPVTFVNRFGLELLKVSQKM
            QKRDKNVHVIVGPGCGSKWAANFADPTTYKPKPNKWDYHGEVNVVIDFY
            GWLPWDDLLDRPLTFTVETGTVFPLARSILITISNTPLEWSSYSAVPAVEALYR
            RITSLVFWKNATPEQSTEEGGQFVTLSPQCEPPEYINY"
            674..679
            complement(704..1018)
            /note="predicted 11.9 kda protein"
            /codon_start=1
            /product="ORF3"
            /protein_id="AAC35311.1"
            /db_xref="PID:g3598811"
            /db_xref="GI:3598811"
            /translation="WVTIPPLVSRMFPVCGFRVCKISSPFAFTTPRPNDVYIGLPI
            TLHFPAHFQKFSQPAETSDKRYRVLLCNGHOTPALQOQTHSSRQVTPLSLRSSSTF
            NK"
            complement(733..912)
            /note="predicted 6.5 kda protein"
            /codon_start=1
            /product="ORF4"
            /protein_id="AAC35312.1"

polyA_signal
CDS
    alignment_scores:
        Quality: 295.00      Length: 59
        Ratio: 5.175        Gaps: 0
        Percent Similarity: 96.610      Percent Identity: 94.915
    alignment_block:
        US-09-209-961-7 x AF055392/rev
        Align seg 1/1 to reverse of: AF055392 from: 1 to: 1768
        1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        912 ATGAGCGTGACATGGTCTTCCAATCACGCTTCGATTTTCCCGCTCAC 863
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        862 TTTCAAAGTTTCAGCAGCCCGCGGAAATTTCTGACAAACGTTACAGGT 813
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        34 yScySPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        812 GCTGCTCTGCACAGGTTCACAGACTCCCGCTCTCCAACAAGGTACTCACA 763
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        51 AlaValAspArgSerLeuArgCysPro 59
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        762 GCAGTAGACAGGTCTACCTCGTGTCTCT 736

BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN
    alignment_scores:
        Quality: 295.00      Length: 59
        Ratio: 5.175        Gaps: 0
        Percent Similarity: 96.610      Percent Identity: 94.915
    alignment_block:
        US-09-209-961-7 x AF055392/rev
        Align seg 1/1 to reverse of: AF055392 from: 1 to: 1768
        1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        912 ATGAGCGTGACATGGTCTTCCAATCACGCTTCGATTTTCCCGCTCAC 863
        ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        862 TTTCAAAGTTTCAGCAGCCCGCGGAAATTTCTGACAAACGTTACAGGT 813
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        34 yScySPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        812 GCTGCTCTGCACAGGTTCACAGACTCCCGCTCTCCAACAAGGTACTCACA 763
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        51 AlaValAspArgSerLeuArgCysPro 59
        ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
        762 GCAGTAGACAGGTCTACCTCGTGTCTCT 736
```

```
seq_name: gb_v1:AF109397

seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
KEYWORDS AF109397.1 GI:4210313
SOURCE bovine circovirus.
ORGANISM bovine circovirus
VIRUSES: ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE 3 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REMARK Sequence update by submitter
COMMENT On Feb 2, 1999 this sequence version replaced gi:4028609.
FEATURES
source
1..1768
/organism="bovine circovirus"
/vifion
/db_xref="taxon:85542"
/note="BCV"
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217"
13..35
/rpt_type=tandem
/rpt_unit=13..18
51..95
/note="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:94210314"
/db_xref="GI:4210314"
/translation="MPSKNGRSGPQHKRWFTLNNPSEDERKKIRELPISLDFYFI
VGEENEGRPHLOGFANFYKQTFNKVYLGAKGTDOONKEYCSKEGN
LLIEGAPRSQGSRLSTAVSLLESGSLVTAEQHPVTFVRNFRGLAELLKVSQM
QRDKNTVHVIVGPGGSKWAANFADPTTYWKPPRNKWDYHGEEVVYIDDFY
GWLPPDDLLCDRLPLVETNGTVFPLRSILTSNOTPLEWYSYTAFAVEALYR
RITSLVFNWKEQSTEEGGQFVTLSPCPPEPPYEINY"
327..332
complement(357..671)
/note="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
/db_xref="PID:94210315"
/db_xref="GI:4210315"
/translation="MVTIPLVSRWFPVCGFRNCKISSPFAATTFRPHNDVYIGLPI
TLLHFPAPHFQSPAEISDRYRVLNCGHOTPALQOQTHSSRQVTLPLSRSTF
NK"
complement(386..565)
/note="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AAD11931.1"
/db_xref="PID:94210316"
/db_xref="GI:4210316"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD11927.1"
/db_xref="PID:94028610"
/db_xref="GI:4028610"
/translation="MVTSLWHLGVVKGANGLLILQPKPHTGNHLETSGGMVTVMKVK
LLMTFMAGCRGMV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD11935.1"
/db_xref="PID:94028618"
/db_xref="GI:4028618"
/translation="MDIDHTVSDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD11937.1"
/db_xref="PID:94028620"
/db_xref="GI:4028620"
/translation="MNNKNHYEVIRKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AAD11932.1"
/db_xref="PID:94028615"
/db_xref="GI:4028615"
/translation="MVFIHLGKGVFKFSELYIHGYTDIVLVVTFVFSAEAE
YVHISRG1"
complement(1022..1027)
complement(1256..1735)
/note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AAD11929.1"
/db_xref="PID:94028612"
/db_xref="GI:4028612"
/translation="MTYPRRYRRRTPRSHLQILRRRPWLHPHRYRWRRKNGI
FNTRLRRTFGYTVKATTVTPSWAVDMRENIDDFVPGGNTNKISIPFEYVIRKVK
VEFWPCSPITQDGRGVSTAVILDDNFVTKATLYDVPVYSSRRTIQQPSYHSR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD11933.1"
/db_xref="PID:94028616"
/db_xref="GI:4028616"
/translation="MASSTPASPAPSDILSRLPQSERPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD11936.1"
/db_xref="PID:94028619"
/db_xref="GI:4028619"
/translation="MSTAQGLTVVVALTVYPKVRERRVLRKMPFFLLQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD11934.1"
/db_xref="PID:94028617"
/db_xref="GI:4028617"
/translation="NAAAGAPSSAVTPPWIRHS"
join(1750..1768,1..13)
stem_loop
```

[illegible]

```

VEFWPCSPITQDGRGVSSAILDDNFVFKATQYDPYVNYSSRHTIPQPFYSHSRY
FTPFPVLDSTIDYFQPNKRNQLWRLQTSRNVHVGIGTAFENSKYDQDYNIRVTMY
VQREFNLKDPPLXP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVVTPKVRERALKMPFFLQF"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRLPQSARPPGR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGVSSSAETPPWIRHS"
join(1750..1768,1..13)
/note="putative; similar to the nonnucleotide motif of
Porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
448 a 362 c 495 g 463 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 295.00      Length: 59
  Ratio: 5.175        Gaps: 0
  Percent Similarity: 96.610      Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF109399/rev ..
Align seg 1/1 to reverse of: AF109399 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
565 ATGACGTGTACATTGGCTCTCCAAATCAGCTTCTGCATTTCCCGCTCAC 516

17 rPheLysSerAlaSerProArgLysPheLeuThrAsnValThrGlyc 34
515 TTTCAAAAGTTCCAGCCAGCCGCGGAAATTTCTGACAAAGTTACAGGT 466

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
465 GCTGCTCTCCAAAGTTCCAGCTCTCCCGCTCTCCAAAGTTACTACACA 416

51 AlaValAspArgSerLeuArgCysPro 59
415 GCAGTAGACAGGTCTACCTCCGTTGTCTCT 389

seq_name: gb_v1:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA VRL 08-JUN-1999
DEFINITION Porcine circovirus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
KEYWORDS porcine circovirus type 2.
SOURCE porcine circovirus type 2.
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and

```

```

Sorden, S.
Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
Unpublished
2 (bases 1 to 1768)
Yoon, K.-J. and Pogranichnyy, R.M.
Direct Submission
Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA
FEATURES
  source
    1..1768
    /organism="porcine circovirus type 2"
    /isolate="ISUVDL 98-15237"
    /specific_host="Sus scrofa"
    /db_xref="taxon:85708"
    /note="isolated from pigs with postweaning multisystemic
    wasting syndrome"
    51..995
    /note="ORF1; 35-36kD"
    /codon_start=1
    /product="unknown"
    /protein_id="AAD37776.1"
    /db_xref="PID:g5007011"
    /db_xref="GI:5007011"
    /translation="MPSKNGRSGPQPKRWVFTLNNPSEDERKKIRELPISLFDYFI
    VGESENGRTPHLQGFANFYKOTFNKVKVTLGARCHIEKAKGTDOONKEYCSKEGN
    LLIECGRPSQGRSDLSSTAVSTLLESGLTVAEQHPVTVFVRNFRGLAELLKVKGM
    QKRWKTNHVIVGPGGKSKWAANFADPETTYWKPPRNKWDGTHGEVVDYDFY
    GWLPWDDLRLCDRYPVTVEKGTGVPFLARSLITLSNOTPLEMYSSTAVPAVEALYR
    RITSLFWKNATEQSTERGQFVTLSPPCPEFFYEINT"
    complement(1034..1735)
    /note="ORF2; 27-28kD"
    /codon_start=1
    /product="unknown"
    /protein_id="AAD37777.1"
    /db_xref="PID:g5007012"
    /db_xref="GI:5007012"
    /translation="MTYPRRYRRRHRPRSHGOILRRPDLVHPHRYFWRKNGI
    ENRLSRFTGYTVATVTVTFSWADMMRFNIDDFVPGGGTKNKISIEFYIRKKVK
    VEFWPCSPITQDGRGVSSAILDDNFVFKATQYDPYVNYSSRHTIPQPFYSHSRY
    FTPFPVLDSTIDYFQPNKRNQLWRLQTSRNVHVGIGTAFENSKYDQDYNIRVTMY
    VQREFNLKDPPLXP"
450 a 362 c 494 g 462 t
BASE COUNT
ORIGIN

alignment_scores:
  Quality: 295.00      Length: 59
  Ratio: 5.175        Gaps: 0
  Percent Similarity: 96.610      Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x AF147751/rev ..
Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768
1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
565 ATGACGTGTACATTGGCTCTCCAAATCAGCTTCTGCATTTCCCGCTCAC 516

17 rPheLysSerAlaSerProArgLysPheLeuThrAsnValThrGlyc 34
515 TTTCAAAAGTTCCAGCCAGCCGCGGAAATTTCTGACAAAGTTACAGGT 466

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
465 GCTGCTCTCCAAAGTTCCAGCTCTCCCGCTCTCCAAAGTTACTACACA 416

51 AlaValAspArgSerLeuArgCysPro 59
415 GCAGTAGACAGGTCTACCTCCGTTGTCTCT 389

```

DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl0111-48285.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-AL.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization

PS of pregnant sows
 CC Claim 14; Fig 2; 48pp; French.
 CC The present sequence represents the nucleotide sequence of PCV isolate
 CC Impl0111-48285. The specification describes a preparation of type II
 CC porcine circovirus (PCV), which is particularly isolated from a lesion,
 CC from a pig with symptoms of PMWS (porcine multisystemic wasting
 CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1767 BP; 448 A; 359 C; 500 G; 460 T;

alignment_scores:

Quality: 295.00 Length: 59
 Ratio: 5.175 Gaps: 0
 Percent similarity: 96.610 Percent Identity: 94.915

alignment_block:

US-09-209-961-7 x X35211/rev ..

Align seg 1/1 to reverse of: X35211 from: 1 to: 1767

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
 ||| |||||
 912 ATGACGTGTACATTAGTCTTCCATCAGCTCTCGCATTTCCGCTCAC 863
 17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
 |||||
 862 TTTCAAAAGTTCAGCCAGCGCGGAAATTTCTGACAAACGTTACAGGT 813
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
 |||||
 812 GCTGCTCTGCAACGGTCCAGACGCTCCGCTCTCCAAACAGGTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59
 |||||
 762 GCAGTAGACAGGTCACTCCGTGTCTCC 736

seq_name: N_Geneseq_36:X35012

seq_documentation_block:

ID X35012 standard; DNA; 1768 BP.

AC X35012;
 DT 01-JUL-1999 (first entry)
 DE Genomic DNA sequence of PCV strain 999PCV.
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
 KW vaccine; ss.
 OS Porcine circovirus
 PN FR2769321-AL.
 PD 09-APR-1999.
 PF 03-OCT-1997; 012382.
 PR 03-OCT-1997; FR-012382.

PA (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYSA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246947/21.
 PT New porcine circovirus from animals with porcine systemic wasting

PT syndrome
 PS Claim 13; Fig 1; 35pp; French.
 CC The present sequence represents the genomic sequence of porcine
 CC circovirus (PCV) strain PCV999. The sequence is obtained from PCV
 CC isolated from a physiological or tissue sample, particularly from
 CC a lesion, from a pig showing symptoms of PMWS (porcine multisystemic
 CC wasting syndrome), or cultured cells, infected with PCV isolated from
 CC such pigs. PCV (attenuated or inactivated), polypeptides derived from
 CC it, and vectors that express these polypeptides are all useful in
 CC vaccines, suitable for administration to adult or young pigs, or to
 CC pregnant sows (for passive immunization of their offspring). DNA
 CC isolated from PCV is used for in vivo or in vitro expression of viral
 CC polypeptides, also as probes or primers for diagnosis in usual
 CC hybridization or amplification assays. These polypeptides may also be
 CC used diagnostically to detect PCV-specific antibodies, while antibodies
 CC raised against the polypeptides can be used to detect antigens, in any
 CC usual immunoassay format.
 SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;

alignment_scores:

Quality: 295.00 Length: 59
 Ratio: 5.175 Gaps: 0
 Percent similarity: 96.610 Percent Identity: 94.915

alignment_block:

US-09-209-961-7 x X35012/rev ..

Align seg 1/1 to reverse of: X35012 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
 ||| |||||
 913 ATGACGTGTACATTAGTCTTCCATCAGCTCTCGCATTTCCGCTCAC 864
 17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
 |||||
 863 TTTCAAAAGTTCAGCCAGCGCGGAAATTTCTGACAAACGTTACAGGT 814
 34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
 |||||
 813 GCTGCTCTGCAACGGTCCAGACGCTCCGCTCTCCAAACAGGTACTCACA 764
 51 AlaValAspArgSerLeuArgCysPro 59
 |||||
 763 GCAGTAGACAGGTCACTCCGTGTCTCT 737

seq_name: N_Geneseq_36:X35213

seq_documentation_block:

ID X35213 standard; DNA; 1768 BP.

AC X35213;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence of PCV isolate Impl010.
 KW PCV isolate; type II porcine circovirus; PCV; PMWS;
 KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
 OS Porcine circovirus.
 PN FR2769322-AL.
 PD 09-APR-1999.
 PF 22-JAN-1998; 000873.
 PR 03-OCT-1997; FR-012382.
 PA (MERI-) MERIAL SAS.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-246948/21.
 PT New type II porcine circovirus, used for, e.g. passive immunization
 PT of pregnant sows

```
PS Claim 14; Fig 4; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Impl010. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 451 A; 361 C; 495 G; 461 T;

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x X35213/rev ..

Align seg 1/1 to reverse of: X35213 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
|||||
912 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCATTTTCCCGCTCAC 863

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
862 TTTCAAAAGTTTCAGCCAGCCCGCGGAAATTTCTGACAAACGTTACAGGTT 813

34 ysCysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
812 GCTGCTCTGCACGGTCCACGACCTCCGCTCTCCAAACGTTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59
|||||
762 GCAGTAGACAGTCACTCGTTGTCTT 736

seq_name: N_Geneseq_36:X35212

seq_documentation_block:
ID X35212 standard; DNA; 1768 BP.
AC X35212;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (corrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Haines D, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 3; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
```

```
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 360 C; 494 G; 461 T;

alignment_scores:
Quality: 295.00 Length: 59
Ratio: 5.175 Gaps: 0
Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
US-09-209-961-7 x X35212/rev ..

Align seg 1/1 to reverse of: X35212 from: 1 to: 1768

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThr 17
|||||
912 ATGACGTGTACATTGGTCTTCCAATCAGCTTCTGCATTTTCCCGCTCAC 863

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
862 TTTCAAAAGTTTCAGCCAGCCCGCGGAAATTTCTGACAAACGTTACAGGTT 813

34 ysCysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
812 GCTGCTCTGCACGGTCCACGACCTCCGCTCTCCAAACGTTACTCACA 763

51 AlaValAspArgSerLeuArgCysPro 59
|||||
762 GCAGTAGACAGTCACTCGTTGTCTT 736

seq_name: N_Geneseq_36:X35214

seq_documentation_block:
ID X35214 standard; DNA; 1768 BP.
AC X35214;
DT 01-JUL-1999 (first entry)
DE Nucleotide sequence of PCV isolate Imp999 (uncorrected version).
KW PCV isolate; type II porcine circovirus; PCV; PMWS;
KW porcine multisystemic wasting syndrome; pig; vaccine; ss.
OS Porcine circovirus.
PN FR2769322-Al.
PD 09-APR-1999.
PF 22-JAN-1998; 000873.
PR 03-OCT-1997; FR-012382.
PA (MERI-) MERIAL SAS.
PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
PI Haines D, Harding J, Hassard L, Meehan B;
DR WPI: 99-246948/21.
PT New type II porcine circovirus, used for, e.g. passive immunization
PT of pregnant sows
PS Claim 14; Fig 6; 48pp; French.
CC The present sequence represents the nucleotide sequence of PCV isolate
CC Imp999. The specification describes a preparation of type II
CC porcine circovirus (PCV), which is particularly isolated from a lesion,
CC from a pig with symptoms of PMWS (porcine multisystemic wasting
CC syndrome). PCV (attenuated or inactivated), polypeptides derived from
CC it, and vectors that express these polypeptides are all useful in
CC vaccines, suitable for administration to adult or young pigs, or to
CC pregnant sows (for passive immunization of their offspring). DNA
CC isolated from PCV is used for in vivo or in vitro expression of viral
CC polypeptides, also as probes or primers for diagnosis in usual
CC hybridization or amplification assays. These polypeptides may also be
CC used diagnostically to detect PCV-specific antibodies, while antibodies
CC raised against the polypeptides can be used to detect antigens, in any
CC usual immunoassay format.
SQ Sequence 1768 BP; 453 A; 359 C; 486 G; 455 T;
```

alignment_scores:
 Quality: 295.00 Length: 59
 Ratio: 5.175 Gaps: 0
 Percent Similarity: 96.610 Percent Identity: 94.915

alignment_block:
 US-09-209-961-7 x X35214/rev ..

Align seg 1/1 to reverse of: X35214 from: 1 to: 1768

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
913 ATGACGTGTACATTGGTCTTCCAAATCAGCTTCTGCATTTTCGCGCTCAC 864

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
863 TTTCAAAAGTTTCAGCCAGCCGCGGAATTTCTGACAAAACGTTACAGGT 814

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
813 GCTGCTCTGCAACGGTCACAGACTCCCGCTCTCCAAACAGGTACTCACA 764

51 AlavalaspargSerLeuArgCysPro 59
|||||
763 GCAGTAGACAGGTGCTCGCTGTGCTCT 737

```

seq_name: N_Geneseq_36:X35013

seq_documentation_block:
 ID X35013 standard; DNA; 1759 BP.
 AC X35013;
 DT 01-JUL-1999 (first entry)
 DE Nucleotide sequence which has homology to PCV sequence.
 KW PCV; strain PCV999; pig; PMWS; porcine multisystemic wasting syndrome;
 KW vaccine; ss.

OS Sus sp.
 FT FR2769321-AL.
 PD 09-APR-1999.
 PF 03-OCT-1997: 012382.
 PR 03-OCT-1997: FR-012382.
 PA (MERI-) MERIAL SAS.
 PA (UYBE-) UNIV QUEENS BELFAST.
 PA (UYGA-) UNIV SASKATCHEWAN.
 PI Allan G, Chappuis GE, Charreyre CE, Clark E, Ellis J,
 PI Haines D, Harding J, Hassard L, Meehan B;
 DR WPI; 99-245947/21.
 PT New porcine circovirus from animals with porcine systemic wasting syndrome

PS Disclosure; Fig 2; 35pp; French.
 CC The specification describes a genomic sequence of porcine circovirus (PCV) strain PCV999. The sequence is obtained from a PCV isolated from a physiological or tissue sample, particularly from a lesion, from a pig showing symptoms of PMWS (porcine multisystemic wasting syndrome), or cultured cells, infected with PCV isolated from such pigs. PCV (attenuated or inactivated), polypeptides derived from it, and vectors that express these polypeptides are all useful in vaccines, suitable for administration to adult or young pigs, or to pregnant sows (for passive immunization of their offspring). DNA isolated from PCV is used for in vivo or in vitro expression of viral polypeptides, also as probes or primers for diagnosis in usual hybridization or amplification assays. These polypeptides may also be used diagnostically to detect PCV-specific antibodies, while antibodies raised against the polypeptides can be used to detect antigens, in any usual immunoassay format. The present sequence appears in the CC specification.

SQ Sequence 1759 BP; 475 A; 510 C; 342 G; 432 T;

alignment_scores:
 Quality: 231.00 Length: 59
 Ratio: 4.529 Gaps: 0
 Percent Similarity: 86.441 Percent Identity: 79.661

alignment_block:
 US-09-209-961-7 x X35013 ..

Align seg 1/1 to: X35013 from: 1 to: 1759

```

1 MetLysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuTh 17
|||||
860 ATGACGTGTACAGCTGTCTTCCAAATCAGCTGCTGCATCTTCCGCTCAC 909

17 rPheLysSerSerAlaSerProArgLysPheLeuThrAsnValThrGlyC 34
|||||
910 TTTCAAAAGTTTCAGCCAGCCGCGGAATTTCTACATACGTACAGGGA 959

34 yScysPheAlaThrValThrArgIleProLeuSerAsnLysValLeuThr 50
|||||
960 ACTGTCGCTACAGTCACCAAGACCCGCTCTCCAAAAGGTACTCACA 1009

51 AlavalaspargSerLeuArgCysPro 59
|||||
1010 GCAGTAGACAGGTGCTGCGCTTCCCTC 1036

```

seq_name: N_Geneseq_36:T84052

seq_documentation_block:
 ID T84052 standard; DNA; 925 BP.
 AC T84052;
 DT 27-AUG-1998 (first entry)
 DE DNA encoding a Staphylococcus aureus protein of unknown function.
 KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
 KW Staphylococcal gene; regulatory element; bacterial gene expression;
 KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
 KW toxic shock syndrome; ss.
 OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT CDS 285..452
 FT /*tag= a

FT WO9730070-AL.
 PD 21-AUG-1997.
 PF 19-FEB-1997: U02318.
 PR 20-FEB-1996; US-011888.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
 PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
 DR WPI; 97-424969/39.
 DR P-PSDB; W28121.
 PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used to isolate antimicrobial compounds, and in vaccines against S. aureus infection

PS Claim 9; Page 846; 989pp; English.

CC The present sequence encodes a Staphylococcus aureus protein of unknown function. The present sequence was isolated from a library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in the construction of ribozymes and antisense sequences to control the expression of staphylococcal genes. The DNA sequence is also useful as a source of regulatory elements for the control of bacterial gene expression. The encoded protein may be used to produce vaccines to enable a host to produce specific antibodies with antibacterial action. These vaccines and antibodies would protect a host against invasion by S. aureus, and conditions relating to Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled skin syndrome, and toxic shock syndrome.

SQ Sequence 925 BP; 327 A; 135 C; 198 G; 261 T;

alignment_scores:
 Quality: 69.50 Length: 44
 Ratio: 2.574 Gaps: 1
 Percent Similarity: 61.364 Percent Identity: 38.636

alignment_block:
 US-09-209-961-7 x T84052/rev ..

Align seg 1/1 to reverse of: T84052 from: 1 to: 925


```

FH Key      Location/Qualifiers
FT CDS      1..741
FT FT       /*tag= a
FT FT       /product= "serine protease"
FT FT       /transl_except= (pos:721..722,aa:Xaa)
FT FT       /note= "Xaa= stop codon"
FT PN       W09740058-Al.
PD 30-OCT-1997.
PF 24-APR-1997; U06121.
PR 04-APR-1997; US-042945.
PR 24-APR-1996; US-639075.
PR 15-NOV-1996; US-749699.
PA (HESK-) HESKA CORP.
PI Frank GR, Gaines PJ, Griewe RB, Hunter SW, Rushlow KE,
PI Silver G, Stiegler GL;
DR WPI; 98-076762/07.
DR P-PSDB; W41974.
PT New flea protease genes and proteins - used in vaccine compositions
PT for the prophylaxis and treatment of flea infestation, especially in
PT cats or dogs
PS Claim 1: Page 259-260; 318pp; English.
CC The present sequence encodes a novel flea serine protease. The
CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
CC protein, as well as the DNA encoding the protein, may all be used in
CC therapeutic compositions to reduce flea protease activity (especially
CC immunoglobulin protease) and so reduce flea infestation, especially in
CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
CC fleas which comprises anti-protease antibodies produced by a host animal
CC in response to administration of the protein. Therapeutic compositions
CC may further comprise a compound that reduces haematophagous ectoparasite
CC burden by a method other than by reducing flea immunoglobulin protease
CC activity. The novel flea DNA encoding the protein can also be used to
CC produce recombinant protein, and fragments of it are used as probes and
CC primers for identification and isolation of related sequences, also as
CC antisense, triplex-forming agents and ribozymes for inhibition of the
CC synthesis of the protein. Ab are also useful for screening expression
CC libraries, to purify the protein and to target cytotoxins to fleas.
SQ Sequence 759 BP; 255 A; 123 C; 172 G; 209 T;

alignment_scores:
  Quality: 62.50      Length: 56
  Ratio: 1.894       Gaps: 2
  Percent similarity: 58.929      Percent Identity: 30.357

alignment_block:
US-09-209-961-7 x V04597/rev ..
Align seg 1/1 to reverse of: V04597 from: 1 to: 759

11 PheCysIlePheProLeuThr.....PheLysSerSerAlaSerPr 24
359 TTTTGACATTTCGCGTGAATTTATTTATTTATTCGGAAGTTTACAGTGC 310
24 OAArgLysPheLeu.....ThrAsnValThrGlyCysCysPheA 37
309 CAATCATTTGCGGCGAGTTGTTTCATCAACTGTCAGTAAATAGTTT 260
37 laThrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAsp 53
259 CTATTCTATATCGTTCACCTACTCTCATCTCATGTTTATTCACGCGGCAACG 210
54 ArgSerLeuArgCysPro 59
209 ACTCAAAACGTTGTGCC 192

seq_name: N_Geneseq_36:Q67959

seq_documentation_block:
ID Q67959 standard; DNA; 627 BP.
AC Q67959;
DT 02-JAN-1995 (first entry)
DE Sequence encoding HBV core-pres1.

```

```

KW Core antigen; recombinant replicable vaccinia virus; hepatitis;
KW prevention; therapy; epitope; ss.
OS Hepatitis B virus
FH Key      Location/Qualifiers
FT CDS      1..627
FT FT       /*tag= a
FT PN       W09412617-A.
PD 09-JUN-1994.
PF 24-NOV-1993; U11474.
PR 25-NOV-1992; US-982211.
PA (ITBI-) INT BIOTECHNOLOGY LAB INC.
PI Bernstine EG, Lewis T, Okeefe RW, Souw PTS;
DR WPI; 94-200247/24.
DR P-PSDB; R55279.
PT Prevention and treatment of hepatitis - using recombinant
PT replicable vaccinia viruses contg. hepatitis B virus surface and
PT core antigen nucleotide sequences
PS Example: Fig 44; 252pp; English.
CC A vector was constructed to allow expression of a core-pres1 fusion
CC polypeptide (called core-pres1* because only a portion of pres1 is
CC terminus of the fusion protein, of AA residues 1-145 or core, a 3 AA
CC spacer (SAC), AAs 1-56 of pres1, and a 4 AA tail (RPTS). To make
CC this fusion vector, a piece of the pres1 region was amplified by PCR
CC from pUEH-04. One primer used was Q67983 which hybridises from bps
CC -22 to -3 relative to the ATG of pres1 and is designed to introduce
CC a BspEI site and to destroy a potential termination codon. A second
CC primer was Q67984 which hybridises to the opposite strand from bps
CC +171 to +152 of the pres1 region and is designed to create an
CC EcoRI site upon amplification. The resulting plasmid, pHTL-26,
CC consists of the modified p7.5 promoter oriented so as to drive
CC expression of the core-pres1* fusion. The sequences are given in
CC Q67959 and R55279.
SQ Sequence 627 BP; 154 A; 157 C; 138 G; 178 T;

alignment_scores:
  Quality: 61.50      Length: 54
  Ratio: 1.809       Gaps: 3
  Percent similarity: 52.963      Percent Identity: 35.185

alignment_block:
US-09-209-961-7 x Q67959 ..
Align seg 1/1 to: Q67959 from: 1 to: 627

9 SerArgPheCysIlePhePro.LeuThrPheLysSerSerAlaSerProA 25
47 TCTCGTTTTCGCTTCGACTTCTTCTTCTCCGT...CAGAGATCTCCTA 93
25 xGlySPheLeuThrAsnValThrGlyCys.....Cys 35
94 GACACCGCCTCAGCTGTGTATCGGGAAGCCTTAGAGTCTCTGTAGGATGC 143
36 PheAlaThrValThrArgIleProLeuSerAsnLysValLeuThrAlaVa 52
144 TTACCTCACCACACCGCAGCTCAGCAGCAGCCATCTCTGTGGGGGAAT 193
52 lAspArgSer 55
194 TGATGACTCT 203

seq_name: N_Geneseq_36:X06815

seq_documentation_block:
ID X06815 standard; mRNA; 1798 BP.
AC X06815;
DT 26-APR-1999 (first entry)
DE Firefly luciferase mRNA.
KW Gene activation; gene expression; cancer; antisense; therapy;
KW beta-galactosidase; luciferase; firefly; ss.
OS Photinus sp.
PN W09858944-Al.

```

PD 30-DEC-1998.
PF 24-JUN-1998; UI3093.
PR 25-JUN-1997; US-050772.
PA (BLAC/) BLACK C A.
PI BLACK CA;
DR WPI; 99-095360/08.
PT New masked targetted expression cassette - useful for regulating
PT gene expression, and preventing neoplastic cell growth
PS Disclosure: Page 28-29; 43pp; English.
CC This is a firefly luciferase full-length mRNA sequence. A new
CC method for activating genes of interest uses a masked targetted
CC expression cassette which expresses a gene product only in the
CC presence of a target molecule. Plasmid pCI-Neo was designed for
CC the production of the sense RNA strand of a targetted expression
CC cassette. Alternative flanking sequences (see X06801-07)
CC corresponding to portions of the firefly luciferase mRNA can be
CC inserted into the multiple cloning sites (see X06800) of the
CC plasmid such that transcription from the T7 promoter yields RNA
CC comprising luciferase segment-Kozak-beta-galactosidase. Sense
CC strand RNA of a masked targetted cassette is produced by in vitro
CC transcription of the construct. Antisense sequences (see X06808-14)
CC corresponding to portions of the target molecule are hybridised to
CC complementary flanking sequences of the sense strand of the
CC targetted cassette. The hybridised mixture is introduced to an in
CC vitro translation mixture containing ribosomes and full-length
CC firefly luciferase RNA. After completion of translation, the
CC mixture is assayed for beta-galactosidase and luciferase
CC activities. Negative luciferase and positive beta-galactosidase
CC activity indicates successful inhibition of the target molecule and
CC successful expression of the gene of interest. The cassette is
CC useful for the treatment of disease and for preventing the
CC proliferation of neoplastic cells. The cassette only releases the
CC antisense strand and expresses a gene product in the presence of a
CC target molecule, so is more potent, reliable, and specific than
CC prior art antisense technology.
SQ sequence 1798 BP; 540 A; 375 C; 423 G; 460 U;

alignment_scores:
Quality: 65.00 Length: 74
Ratio: 1.912 Gaps: 6
Percent Similarity: 45.946 Percent Identity: 33.784

alignment_block:
US-09-209-961-7 x X06815/rev ..

Align seg 1/1 to reverse of: X06815 from: 1 to: 1798

```
11 PheCysIlePheProLeuThrPheLysSerAlaSerProArgLysPh 27
|||||
471 TTTTGC.....ACGTCAAAATTTTTCGAACCC.....TT 440

27 eLeu...ThrAsnValThr...GlyCys..... 34
|||||
439 TTGGAAACAACACTACGGTAGCGTGGGAATGTTCACTGTTGACCA 390

35 .....CysPheAlaThrValThrArgIleProLeu 44
|||||
389 ATTCACTGTCATTATAATGTGCTTCGGGGCGCAACTGCAACTCCGATA 340

45 SerAsn.....LysValLeuThrAl 51
|||||
339 AATACGGCGCCCAACACCGCATAAAGATTGAAGAGAGTTTTCACCTGC 290

51 aValAspArgSerLeuArgCys 58
|||||
289 ATACGACGATTCTGTGATTGT 268
```

seq_name: N_Geneseq_36:V74820

seq_documentation_block:

ID V74820 standard; DNA; 1373 BP.
AC V74820;

DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #509.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
OS Staphylococcus aureus;
PH Key Location/Qualifiers
FT misc_feature 961..1020
FT /*tag=a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence"
PN EP-786519-A2.
PD 30-JUL-1997. 100117.
PF 07-JAN-1997; US-009861.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 1436-1437; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 1373 BP; 363 A; 258 C; 183 G; 506 T;

alignment_scores:
Quality: 63.50 Length: 32
Ratio: 3.175 Gaps: 1
Percent Similarity: 62.500 Percent Identity: 46.875

alignment_block:
US-09-209-961-7 x V74820 ..

Align seg 1/1 to: V74820 from: 1 to: 1373

```
15 ProLeuThrPheLysSerAlaSerProArgLysPheLeuThrAsnVa 31
|||||
505 CCATATATCTTTTAGTTCATCGATAGTTTGTAGCTAAATTTGCTACTCGAAT 554

31 LhrGlyCysCys.....PheAlaThrVal 39
|||||
555 AACTGTCATCATCTTCAATTCACCTGTTTCAGGCTTTTGCAACTGTT 600
```

seq_name: N_Geneseq_36:T98716

seq_documentation_block:

ID T98716 standard; DNA; 1168 BP.

AC T98716;

DT 09-NOV-1998 (first entry)

DE DNA encoding a S. pneumoniae protein of unknown function.

KW Streptococcus pneumoniae protein; genetic immunisation; antagonist;

OM of: US-09-209-961-7 to: Issued_Patents_NA:* out_format : pfs
Date: Dec 31, 1999 12:26 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 Compugen Ltd.

Command line parameters:
-MODEL=frame1.p2n.model -DRV=xlp
-O/cgn2_1/USPTO.spool/US09209961/runat.22121999_101043_17921/app_query.fasta.1
-DB=Issued_Patents_NA -OEMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=escore
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=stat -USER=US09209961
-NCPU=6 -TCPU=3 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-209-961-7
Query length: 59
Database: Issued_Patents_NA*
Database sequences: 195662
Database length: 53930012
Search time (sec): 88.010000

```
score_list:
sequence      Strd Orig      ZScore      EScore Len      Documentation
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-7 + 62.50 121.79 21.84 2330 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-470-179-139 + 54.50 120.14 27.00 423 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-125-462-8 + 54.00 120.01 27.44 381 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-891-848-8 + 54.00 120.01 27.44 381 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-1 + 62.50 118.72 32.39 3115 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-17 + 58.00 118.71 32.44 1094 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-30 + 58.00 118.71 32.44 1094 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-17 + 58.00 118.71 32.44 1094 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-30 + 58.00 118.71 32.44 1094 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-5 + 62.50 118.27 34.32 3251 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-589-711-1 + 61.50 117.96 35.72 2653 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-739-485-6 + 58.50 117.82 36.33 1336 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-125-462-9 + 54.00 116.32 44.04 540 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-891-848-9 + 54.00 116.32 44.04 540 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-009-261-4 + 53.50 116.08 45.45 492 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-451-715A-7 - 60.50 114.29 57.17 2973 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-867-129-1 - 58.50 111.08 86.33 2529 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-325-071-66 - 57.50 110.35 94.81 2147 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-791-213D-14 + 47.00 109.08 111.56 210 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-293-150A-14 + 47.00 109.08 111.56 210 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-7 + 49.00 108.78 115.86 344 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-7 + 49.00 108.78 115.86 344 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-057-971-7 + 49.00 108.78 115.86 344 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-07-743-518-25 + 50.50 108.11 126.36 520 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-450-384-7 - 43.00 107.84 130.71 93 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-309-604-1 - 48.50 107.61 134.61 342 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-450-384-6 - 43.00 107.18 142.27 99 !
/cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-392-546G-3 - 50.50 106.30 159.35 617 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-093-144-6 - 49.00 106.08 163.77 444 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-755-462-1 - 49.50 106.06 184.30 500 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-199-317-1 - 49.50 106.06 184.30 500 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-450-065-1 - 50.50 106.01 165.33 634 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-450-595-1 - 50.50 106.01 165.33 634 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-840-683-1 + 53.00 105.49 176.76 1192 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-27 + 48.00 105.30 181.19 379 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-27 + 48.00 105.30 181.19 379 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-057-971-27 + 48.00 105.30 181.19 379 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-068-395A-2 + 54.50 105.09 186.01 1755 !
/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-464-365-2 + 54.50 105.09 186.01 1755 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-791-213D-88 + 47.00 104.86 191.68 313 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-972-387-1 + 47.00 104.86 191.68 313 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-431-412-1 + 47.00 104.86 191.68 313 !
```

/cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-057-971-1 + 47.00 104.86 191.68 313 !
/cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-293-150A-88 + 47.00 104.86 191.68 313 !
/cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-370-975B-4 + 48.00 104.67 196.27 402

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-7

seq_documentation_block:

; Sequence 7, Application US/09132619B
; Patent No. 5958697
; GENERAL INFORMATION:
; APPLICANT: Shan, Bei
; TITLE OF INVENTION: CYP7 Promoter-Binding Factors
; FILE REFERENCE: T97-013
; CURRENT APPLICATION NUMBER: US/09/132,619B
; EARLIER FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 60/067,708
; EARLIER FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 2330
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (363)..(1862)
US-09-132-619-7

alignment_scores:

Quality: 62.50 Length: 32
Ratio: 2.717 Gaps: 1
Percent Similarity: 71.875 Percent Identity: 43.750

alignment_block:

US-09-209-961-7 x US-09-132-619-7 ..

Align seg 1/1 to: US-09-132-619-7 from: 1 to: 2330

14 PheProLeuThrPhelySerSerAlaSerProArgLys...PheLeuTh 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
868 TTTCTCTGCAATTCAAACATCCATCTGCTCCCAAGGCGCTACCTCTG 917

29 rAsnValThrGlyCysCysPheAlaThrValThrArgileProLeu 44
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

918 AACCATGCTGCTGCTGCTCTACAGACTATGACAGAGTCCCTTTG 963

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-470-179-139

seq_documentation_block:

; Sequence 139, Application US/08470179
; Patent No. 5645994
; GENERAL INFORMATION:
; APPLICANT: Huang Ph.D, Wai Mun
; TITLE OF INVENTION: Method and Compositions for
; IDENTIFICATION OF SPECIES IN A SAMPLE
; NUMBER OF SEQUENCES: 207
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Trask, Britt and Rossa
; STREET: P.O. Box 2550
; CITY: Salt Lake City
; STATE: Utah
; COUNTRY: USA
; ZIP: 84110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/470,179
; FILING DATE:


```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sweigert Ph.D, Susan E.
; REGISTRATION NUMBER: 36,289
; REFERENCE/DOCKET NUMBER: 2601
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 801-532-1922
; TELEFAX: 801-531-9168
; INFORMATION FOR SEQ ID NO: 139:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 423 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: not relevant
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Leptospira borgpetersenii
; US-08-470-179-139

alignment_scores:
  Quality: 54.50      Length: 64
  Ratio: 1.652       Gaps: 1
  Percent Similarity: 51.562   Percent Identity: 21.875

alignment_block:
  US-09-209-961-7 x US-08-470-179-139 ..
  Align seg 1/1 to: US-08-470-179-139 from: 1 to: 423

11 PheCysIlePhePro..... 15
164 TTCAGGATTTTCCTCCGCGTTCCTCATGATGAGCGAAGAAATTC 213
16 .....LeuThrPheLysSerSerAlaSerProArgLysP 27
214 GGTTCTATAGCGCGGATACCCCTGCGCTTATCGGTACACCGAAGCAAG 263
27 heLeuThrAsnValThrGlyCysCysPheAlaThrValThrArgIlePro 43
264 ACTGTAAAAAGTCGCGAAGAAATGTTACGGCACATTGAAAAAGACCG 313
44 LeuSerAsnLysValLeuThrAlaValAspArgSerLeuArg 57
314 TTAGTTCTCGCCTAATACGATGATACAAACAAACACCGG 355

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-125-462-8

seq_documentation_block:
; Sequence 8, Application US/08125462
; Patent No. 5840840
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Youle, Richard J.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Nicholls, Peter J.
; TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/125,462
; FILING DATE: 22-SEP-1993
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen L.
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 15280-110-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 543-9600
; TELEFAX: (415) 543-5043
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1-381
; OTHER INFORMATION: /standard_name= "Human RNase Gene"
; US-08-125-462-8

alignment_scores:
  Quality: 54.00      Length: 80
  Ratio: 1.317       Gaps: 4
  Percent Similarity: 51.250   Percent Identity: 28.750

alignment_block:
  US-09-209-961-7 x US-08-125-462-8 ..
  Align seg 1/1 to: US-08-125-462-8 from: 1 to: 381

2 LysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThrPh 18
125 CGGTAAACACATTCGTTTCATGAACCGTGG.....TTGACGTT 162
18 eLysSerSerAlaSerProArgLysPheLeu..... 28
163 CAGAACGTTTCCTCCAGGAAAAAGTTACTTGCAGAAACGTCAGGGTAA 212
29 .....ThrAsnValThrGlyCysCysPheAla..... 37
213 CTGCTACAAATCACTTCTTATGATATCATCATGACTGCCGCTGTGACTA 262
38 .....ThrValThrArgIleProLeuSer..... 45
263 ACGGTTCTCGTTACCGGAACGTCGCTTACCGTACTTCTCCGAAAGACGT 312
46 ....AsnLysValLeuThrAlaValAspArgSerLeuArg 57
313 CATATCATCGTTGTTGCGAAGGTTCTCCGTACGTTCGCG 352

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-891-848-8

seq_documentation_block:
; Sequence 8, Application US/08891848
; Patent No. 5955073
; GENERAL INFORMATION:
; APPLICANT: Rybak, Susanna M.
; APPLICANT: Youle, Richard J.
; APPLICANT: Newton, Dianne L.
; APPLICANT: Nicholls, Peter J.
; TITLE OF INVENTION: Selective RNase Cytotoxic Reagents
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
```

```

;
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/891,848
; FILING DATE: 04-FEB-1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/125,462
; FILING DATE: 22-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/014,082
; FILING DATE: 04-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/779,195
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/510,696
; FILING DATE: 20-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Weber, Ellen Lauver
; REGISTRATION NUMBER: 32,762
; REFERENCE/DOCKET NUMBER: 015280-110310US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 576-0200
; TELEFAX: (415) 576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 381 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: -
; LOCATION: 1..381
; OTHER INFORMATION: /note= "human ribonuclease (RNase)"
; US-08-891-848-8
;
alignment_scores:
  Quality: 54.00      Length: 80
  Ratio: 1.317       Gaps: 4
  Percent Similarity: 51.250      Percent Identity: 28.750
;
alignment_block:
  US-09-209-961-7 x US-08-891-848-8 ..
;
Align seg 1/1 to: US-08-891-848-8 from: 1 to: 381
;
2 LysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThrPh 18
... ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
125 CGGTACACATTCGTCATCAACCGCTGG.....TTGACGTT 162
;
18 eLysSerSerAlaSerProArgLysPheLeu..... 28
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
163 CAGAACGTTTCTCCAGGAAAGTACTTGCACAAACGCTCAGGTAA 212
;
29 .....ThrAsnValThrGlyCysPheAla..... 37
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
213 CTGTACAAATCTAATCTTCTATGCATATCACTGACTGCGCTGACTA 262
;
38 .....ThrValThrArgIleProLeuSer..... 45
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
263 ACGTTCTCGTACCCGAACTCCGCTACCTACTTCTCCGAAAGACGT 312
;
46 ....AsnLysValLeuThrAlaValAspArgSerLeuArg 57
..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
313 CATATCATCGTTGCTGCGAAGGTTCTCCGTCGTTCCGG 352
;
seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-09-132-619-1
seq_documentation_block:
; Sequence 1, Application US/09132619B
; Patent No. 5958697
; GENERAL INFORMATION:
; APPLICANT: Nitta, Masahiro
; TITLE OF INVENTION: CYP7 Promoter-Binding Factors
; FILE REFERENCE: T97-013
; CURRENT APPLICATION NUMBER: US/09/132,619B
; CURRENT FILING DATE: 1998-08-11
; EARLIER APPLICATION NUMBER: 60/067,708
; EARLIER FILING DATE: 1997-12-08
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 3115
; TYPE: DNA
; ORGANISM: human
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (210)..(1694)
; US-09-132-619-1
;
alignment_scores:
  Quality: 62.50      Length: 32
  Ratio: 2.717       Gaps: 1
  Percent Similarity: 71.875      Percent Identity: 43.750
;
alignment_block:
  US-09-209-961-7 x US-09-132-619-1 ..
;
Align seg 1/1 to: US-09-132-619-1 from: 1 to: 3115
;
14 PheProLeuThrPheLysSerSerAlaSerProArgLys...PheLeuTh 29
||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
697 TTCTCTGCAATTCAAAATCCACTTGCCTCCAAAGGCTACCTCTG 746
;
29 rAsnValThrGlyCysPheAlaThrValThrArgIleProLeu 44
..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
747 AACCATGTCCTTCCCTCTACAGACATATGACAGAGTCCCTTG 792
;
seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-431-080-17
seq_documentation_block:
; Sequence 17, Application US/08431080
; Patent No. 5698686
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,080
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
;

```

```
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-431-080-17

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
Percent Similarity: 66.667 Percent Identity: 44.444

alignment_block:
US-09-209-961-7 x US-08-431-080-17 ..
Align seg 1/1 to: US-08-431-080-17 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
||||:|||||
70 TTCTCCTCGATTGTTGTTATCATCTTCCACACACTTTCCTGGGCTC 119

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAla 38
||||:|||||
120 CACT.....TTTTGATTGGTGTTCCTTGAATTGTTATCTGA 157

38 hrValThr 40
||||:|||||
158 CTCCTCACT 165

seq_name: /cgn2_6/ptodata/1/lna/5B_COMB.seq:US-08-431-080-30

seq_documentation_block:
; Sequence 30, Application US/08431080
; Patent No. 5698686
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,080
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
```

```
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-431-080-30

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
Percent Similarity: 66.667 Percent Identity: 44.444

alignment_block:
US-09-209-961-7 x US-08-431-080-30 rev ..
Align seg 1/1 to reverse of: US-08-431-080-30 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
||||:|||||
1025 TTCTCCTCGATTGTTGTTATCATCTTCCACACACTTTCCTGGGCTC 976

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAla 38
||||:|||||
975 CACT.....TTTTGATTGGTGTTCCTTGAATTGTTATCTGA 938

38 hrValThr 40
||||:|||||
937 CTCCTCACT 930

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-938-534-17

seq_documentation_block:
; Sequence 17, Application US/08938534
; Patent No. 5916752
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; APPLICANT: Singer, Miriam S.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
```

```
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-938-534-17

alignment_scores:
  Quality: 58.00      Length: 36
  Ratio: 2.417       Gaps: 2
Percent similarity: 66.667 Percent Identity: 44.444

alignment_block:
US-09-209-961-7 x US-08-938-534-17 ..
Align seg 1/1 to: US-08-938-534-17 from: 1 to: 1094

7 PheGlnSerArgPheCys.....IlePheProLeuThrPheLysSerSe 21
||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
1025 TTCCTCTCGATTTGTGTTATCATTTCTCCACAACTTTCTGGGCTC 976

21 rAlaserProArgLysPheLeuThrAsnValThrGlyCysPheAla 38
||||:||||:||||:||||:||||:||||:||||:||||:||||:
975 CACT.....TTTTGATTGGTGTGCTTGAATTGTTATCTGA 938

38 hrValThr 40
||||:||||
937 CTCTCACT 930

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-938-534-30
seq_documentation_block:
; Sequence 30, Application US/08938534
; Patent No. 5916752
; GENERAL INFORMATION:
; APPLICANT: Gottschling, Daniel E.
; TITLE OF INVENTION: Telomerase Compositions and Methods
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TEXAS
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/938,534
; FILING DATE: 26-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/431,080
; FILING DATE:
; APPLICATION NUMBER: SN 08/326,781
; FILING DATE: October 20, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:155/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 30:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1094 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-938-534-30

alignment_scores:
  Quality: 62.50      Length: 32
  Ratio: 2.717       Gaps: 1
Percent similarity: 71.875 Percent Identity: 43.750

alignment_block:
US-09-209-961-7 x US-08-132-619-5 ..
Align seg 1/1 to: US-08-132-619-5 from: 1 to: 3251

14 PheProLeuThrPheLysSerSerAlaSerProArgLys....PheLeu 29
||||:||||:||||:||||:||||:||||:||||:||||:||||:
833 TTCCTCTGCAATTCAAAACATCCACTCTGCCTCCAAAGGCCCTACCT 882

29 rAsnValThrGlyCysPheAlaThrValThrArgIleProLeu 44
||||:||||:||||:||||:||||:||||:||||:||||:||||:
883 AACCATGCTGCTTGCCTCTACAGACTATGACAGAGTCCCTTTG 928
```


Quality: 54.00 Length: 80
Ratio: 1.317 Gaps: 4
Percent Similarity: 51.250 Percent Identity: 28.750
alignment_block:
US-09-209-961-7 x US-08-891-848-9 ..
Align seg 1/1 to: US-08-891-848-9 from: 1 to: 540
2 LysCysThrLeuValPheGlnSerArgPheCysIlePheProLeuThrPh 18
::: ||||| |||||:|||||: |||||
125 CGGTTAACACTTCGTTCAATCAACCGTGG.....TTGACGTT 162
18 eLysSerSerAlaSerProArgLysPheLeu..... 28
|:||||| ||||| :|||:|||||
163 CAGAACGTTGCTTCAGGAAAGATTACTTGCAAAACGGTCAGGTA 212
29ThrAsnValThrGlyCysPheAla..... 37
|||||:||||| |||||:|||||
213 CTGCTACAAATCTAACTCTTCTATGCATATACACTGACTGCGCTGACTA 262
38ThrValThrArgIleProLeuSer..... 45
|||||:||||| |||||:|||||
263 ACGGTTCTGCTTACCGCAAGCTGCGCTTACCGTACTTCTCCGAAAGACGT 312
46AsnValLeuThrAlaValAspArgSerLeuArg 57
:|||||:||||| |||||:|||||
313 CATATCATGCTGCTGCGAAGGTTCTCCGTACGTTCCGG 352

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-08-009-261-4

seq_documentation_block:
; Sequence 4, Application US/08009261
; Patent No. 5350672
; GENERAL INFORMATION:
; APPLICANT: Oberst, Richard D.
; APPLICANT: Gwaltney, Sharon M.
; APPLICANT: Hays, Michael P.
; TITLE OF INVENTION: Specific DNA Primers and Method to Use
; NUMBER OF INVENTIONS: 4
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Tilton, Fallon, Lungmus and Chestnut
; STREET: 100 S. Wacker Drive, Suite 960
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606-4002
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/009,261
; FILING DATE: 19930122
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fentress, Susan B.
; REGISTRATION NUMBER: 31,327
; REFERENCE/DOCKET NUMBER: 92061A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)456-8000
; TELEFAX: (312)456-7776
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
US-08-009-261-4

alignment_scores:
Quality: 53.50 Length: 67
Ratio: 1.672 Gaps: 3
Percent Similarity: 47.761 Percent Identity: 29.851
alignment_block:
US-09-209-961-7 x US-08-009-261-4 ..
Align seg 1/1 to: US-08-009-261-4 from: 1 to: 492
1 MetLysCysThrLeuValPheGlnSerArgPheCys.....IlePh 14
:||||| :||||| |||||:|||||
90 CTACGAAAGTGTTCTTCATCATCAATCTAGGTATCTGACATTAGGATTCTT 139
14 eProLeuThrPhe..... 18
||||| :|||||
140 CCCTCAGAAATTATTCTCGATAAATTCGACTCCCTTCACTGTGCATACCAAT 189
19LysSerSerAlaSerProArgLysPheLeuThrAsn 30
:|||||:||||| |||||:|||||
190 TCTGAGAACTTCTCTCTGAATCTATATCTACACGCAATTCCTGTAAACAT 239
31 ValThrGlyCysPheAlaThrValThrArgIleProLeuSerAsnLy 47
||||| ||||| :|||||:|||||
240 ...ACAGGAGTTTGTCTCTTCATGACACACGACTTCCTCTTATGAATGA 286
47 S 47
287 G 287

241 CTTTACCAAGTTTCATCTTACTGGAGGGTCTTGAATAACAGATCGA 290

55SerLeuArgCysPro 59
||||:|||||

291 CGGGCGCGCGGGTGGCGAATAACTCCAGATGATCCCGC 335

seq_name: gb_gss7:AQ503361

seq_documentation_block: 561 bp DNA GSS 28-APR-1999
LOCUS AQ503361
DEFINITION V57C10 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', genomic survey sequence.
ACCESSION AQ503361
NID 94705907
VERSION AQ503361.1 GI:4705907
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 561)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desEstages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)

TITLE
JOURNAL
COMMENT

Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumare@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTCTTCTTTGGAAGTAC
Class: transposon-tagged.

FEATURES
source
1..561
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 151 a 142 c 115 g 150 t 3 others
ORIGIN

alignment_scores:
Quality: 71.00 Length: 51
Ratio: 2.290 Gaps: 2
Percent Similarity: 60.784 Percent Identity: 39.216

alignment_block:
US-09-209-961-7 x AQ503361/rev ..

Align seg 1/1 to reverse of: AQ503361 from: 1 to: 561

9 SerArgPheCysIlePheProLeuThrPhe.....LysSerSe 21
||| ||| ||||| |||

458 AGTGTCTTTAGTGAATTCCTCCCTGCGTTTCGTAACATACAGATGG 409

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAlat 38
:||||:|||||: :|||||

408 GGCAGACCCACGTCGA.....NACAACGTCGGCATTGCTGTTTCGCA 365

38 hrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAspArg 54
|||:||||: |||:|||||

364 CTTTACCAAGTTTCATCTTACTGGAGGGTCTTGAATAACAGATCAG 315

55 Ser 55
::

314 ACG 312

seq_name: gb_gss7:AQ502889

seq_documentation_block: 505 bp DNA GSS 28-APR-1999
LOCUS AQ502889
DEFINITION V48G6 mTn-3xHA/lacZ Insertion Library Saccharomyces cerevisiae
genomic 5', genomic survey sequence.

ACCESSION AQ502889
NID 94705435
VERSION AQ502889.1 GI:4705435
KEYWORDS GSS.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.

REFERENCE 1 (bases 1 to 505)
AUTHORS Ross-Macdonald, P., Roemer, T., Coelho, P.S.R., Agarwal, S., Kumar, A.,
desEstages, S.A., Cheung, K.-H., Sheehan, A., Symoniatidis, D., Jansen, R.,
Umansky, L., Heidtman, M., Nelson, K., Iwasaki, H., Kanada, D., Lugo, R.,
Hager, K., Miller, P., Roeder, G.S. and Snyder, M.
Large-Scale Analysis of the Yeast Genome by Transposon Tagging and
Gene Disruption
Unpublished (1999)

TITLE
JOURNAL
COMMENT

Contact: Kumar A
Michael Snyder, Dept. of Mol. Cell. and Dev. Biology
Yale University
P.O. Box 208103, New Haven, CT 06520-8103, USA
Tel: 203 432 9949
Fax: 203 432 6161
Email: anuj.kumare@yale.edu
te of mTn-3xHA/lacZ insertion.
Seq primer: GGCCTCTTCTTTGGAAGTAC
Class: transposon-tagged.

FEATURES
source
1..505
Location/Qualifiers
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/clone_lib="mTn-3xHA/lacZ Insertion Library"
/lab_host="E. coli"
/note="Vector: pHS56-Sal; A yeast genomic DNA library
(lacking mitochondrial DNA) was prepared in pHS56-Sal;
genomic DNA was size-fractionated (DNA of roughly 2-3 kb
in length) prior to cloning. This library was
subsequently mutagenized with a mTn-3xHA/lacZ
minitransposon containing lacZ, URA3, and tet resistance."

BASE COUNT 115 a 124 c 90 g 176 t
ORIGIN

alignment_scores:
Quality: 70.00 Length: 52
Ratio: 2.258 Gaps: 2
Percent Similarity: 59.615 Percent Identity: 38.462

alignment_block:
US-09-209-961-7 x AQ502889 ..

Align seg 1/1 to: AQ502889 from: 1 to: 505

9 SerArgPheCysIlePheProLeuThrPhe.....LysSerSe 21
||| ||| ||||| |||

60 AGTGTCTTTAGTGAATTCCTCCCTGCGTTTCGTAACATACAGATGG 109

21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheAlat 38
:||||:|||||: :|||||

Align seq 1/1 to: A0491934 from: 1 to: 816

alignment_block:
US-09-209-961-7 x AQ491969

Align seg 1/1 to: A0491969 from: 1 to: 837

```

9 SerArgPheCysIlePheProLeuThrPhe.....LysSerse 21
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
102 AGTCTTTAGTTCGATTCCTCCCTCGTTCTGTCGAACCATAGATGG 151
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
21 rAlaSerProArgLysPheLeuThrAsnValThrGlyCysCysPheLat 38
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
152 GCGACGCCACGTCGA.....AACACGTCGGCACCTGCTGTTTCCGCA 195
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
38 hrValThrArgIleProLeuSerAsnLysValLeuThrAlaValAspArg 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 CTTTACCAAGTTCATCTTACTGGAGGCTCTTGAAACACGATCAG 245
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
55 Ser 55
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
246 ACG 248
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: gb_est8:C07961

```

seq_documentation_block: 300 bp mRNA EST 12-SEP-1996
LOCUS C07961 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
DEFINITION Caenorhabditis elegans cDNA clone yk172ell 3', mRNA sequence.
ACCESSION C07961
NID 91533032
VERSION C07961.1 GI:1533032
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motoshashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:637700.

```

```

FEATURES
    source
        1..300
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk172ell"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT 75 a 70 c 77 g 78 t
ORIGIN

```

```

alignment_scores:
    Quality: 66.50 Length: 42
    Ratio: 2.293 Gaps: 1
    Percent Similarity: 69.048 Percent Identity: 35.714
alignment_block:
    US-09-209-961-7 x C07961/rev

```

Align seg 1/1 to reverse of: C07961 from: 1 to: 300

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16

```

   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
194 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGTGATGCCAAG 145
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
144 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCTCATCTC 95
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

33 lyCysCysPheAlaThrValThrArg 41

```

   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
94 GCGCATGCTTCTCGGACATTTCCGGC 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

seq_name: gb_est16:C36594

```

seq_documentation_block: 300 bp mRNA EST 10-SEP-1997
LOCUS C36594 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite embryo
DEFINITION Caenorhabditis elegans cDNA clone yk468d4 3', mRNA sequence.
ACCESSION C36594
NID 92372735
VERSION C36594.1 GI:2372735
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE 1 (bases 1 to 300)
AUTHORS Kohara,Y., Motoshashi,T., Tabara,H., Watanabe,H., Sugimoto,A.,
Sano,M., Miyata,A. and Nishigaki,A.
TITLE Expression map of the C.elegans genome
JOURNAL Unpublished (1996)
COMMENT On Dec 30, 1996 this sequence version replaced gi:1529573.

```

```

FEATURES
    source
        1..300
            /organism="Caenorhabditis elegans"
            /strain="N2"
            /db_xref="taxon:6239"
            /clone="yk468d4"
            /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
            hermaphrodite embryo"
            /sex="hermaphrodite"
            /dev_stage="embryo"
BASE COUNT 75 a 69 c 78 g 78 t
ORIGIN

```

```

alignment_scores:
    Quality: 66.50 Length: 42
    Ratio: 2.293 Gaps: 1
    Percent Similarity: 69.048 Percent Identity: 35.714
alignment_block:
    US-09-209-961-7 x C36594/rev

```

Align seg 1/1 to reverse of: C36594 from: 1 to: 300

```

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
195 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGTGATGCCAAG 146
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33

```

   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
145 ATCTATCAGATCTACGAGGGAACCTCGCAATTCAGCGCATGTCTCATCTC 96
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

33 lyCysCysPheAlaThrValThrArg 41

```

   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
95 GCGCATGCTTCTCGGACATTTCCGGC 70
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and Nomoto, H.
 Expressed genes in C.elegans
 Unpublished (1999)
 On Jun 5, 1998 this sequence version replaced gi:3189129.

Contact: Yuji Kohara

Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES

1..300
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
 hermaphrodite-embryo"
 /sex="hermaphrodite"
 /dev_stage="embryo"
 /dev_stage="embryo"
 76 a 66 c 78 g 80 t

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 66.50 Length: 42
 Ratio: 2.293 Gaps: 1
 Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:

US-09-209-961-7 x AV182172/rev ..

Align seg 1/1 to reverse of: AV182172 from: 1 to: 300

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16

209 ATGGCTCAACTCTGAGTATCGGAGAACTGATCGGTGATGCCAAG 160

16 utrPhelySerSerAlaSerProArgLysPheLeuThrAsnValThrG 33

159 ATCTATCAGATCTACGAGGAACTCGCAATTCAGCGCATGTCATCTC 110

33 lyCysCysPheAlaThrValThrArg 41

109 GCGCATGCTTCTCGGACATTTCCGCGC 84

seq_name: gb_est36:AV182309

seq_documentation_block:

LOCUS AV182309 300 bp mRNA EST 21-JUL-1999
 DEFINITION AV182309 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
 embryo Caenorhabditis elegans cDNA clone yk636a12 3', mRNA
 sequence.

ACCESSION AV182309

NID 95562210

VERSION AV182309.1 GI:5562210

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 300)

AUTHORS Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuiki, H.,

Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A.,

Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and

Nomoto, H.

Expressed genes in C.elegans

Unpublished (1999)

On Jun 5, 1998 this sequence version replaced gi:3189269.

Contact: Yuji Kohara

Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855
 Email: ykohara@lab.nig.ac.jp.

FEATURES

1..300
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone_lib="Yuji Kohara unpublished cDNA:Strain N2
 hermaphrodite-embryo"
 /sex="hermaphrodite"
 /dev_stage="embryo"
 /dev_stage="embryo"
 76 a 64 c 75 g 71 t 14 others

BASE COUNT
 ORIGIN

alignment_scores:

Quality: 66.50 Length: 42
 Ratio: 2.293 Gaps: 1
 Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:

US-09-209-961-7 x AV182309/rev ..

Align seg 1/1 to reverse of: AV182309 from: 1 to: 300

1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProLe 16

211 ATGGCTCAACTCTGAGTATCGGAGAACTGATCGGTGATGCCAAG 162

16 utrPhelySerSerAlaSerProArgLysPheLeuThrAsnValThrG 33

161 ATCTATCAGATCTACGAGGAACTCGCAATTCAGCGCATGTCATCTC 112

33 lyCysCysPheAlaThrValThrArg 41

111 GCGCATGCTTCTCGGACATTTCCGCGC 86

seq_name: gb_est36:AV184335

seq_documentation_block:

LOCUS AV184335 300 bp mRNA EST 21-JUL-1999
 DEFINITION AV184335 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
 embryo Caenorhabditis elegans cDNA clone yk663b4 3', mRNA sequence.

ACCESSION AV184335

NID 95564236

VERSION AV184335.1 GI:5564236

KEYWORDS EST.

SOURCE Caenorhabditis elegans.

ORGANISM Caenorhabditis elegans

REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

1 (bases 1 to 300)

AUTHORS Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Mitsuiki, H.,

Nishigaki, A., Motohashi, T., Zeng, Q., Watanabe, H., Sugimoto, A.,

Sano, M., Miyata, A., Mitani, Y., Iida, K., Uesugi, H., Sugiyama, Y. and

Nomoto, H.

Expressed genes in C.elegans

Unpublished (1999)

On Mar 10, 1998 this sequence version replaced gi:2948315.

TITLE

JOURNAL

COMMENT

Contact: Yuji Kohara

Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 81-559-81-6854
 Fax: 81-559-81-6855

```

FEATURES
Source
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/dev_stage="embryo"
BASE COUNT 74 a 69 c 80 g 77 t
ORIGIN

alignment_scores:
Quality: 66.50 Length: 42
Ratio: 2.293 Gaps: 1
Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:
US-09-209-961-7 x AV184335/rev ..
Align seg 1/1 to reverse of: AV184335 from: 1 to: 300
1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProle 16
||||| ||||| ::||| ||| |||||::|||
202 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGGTGATGCCAAG 153
16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33
:::::|||||: ::||| |||||::|
152 ATCTATCATGCTACGAGGAGACCTCGCAATTCAGCGCATGTCATCTC 103
33 lyCysCysPheAlaThrValThrArg 41
:::::|||||: ::||| |||||::|
102 GCGCATGCTCTCGGACATTCGCGC 77

seq_name: gb_est36:AV185677

seq_documentation_block:
LOCUS AV185677 300 bp mRNA EST 21-JUL-1999
DEFINITION AV185677 Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk681h6 3', mRNA sequence.
ACCESSION AV185677
NID 95565578
VERSION AV185677.1 GI:5565578
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara,Y., Shin-i,T., Thierry-Mieg,J., Thierry-Mieg,D., Mitsuki,H.,
Nishigaki,A., Motohashi,T., Zeng,Q., Watanabe,H., Sugimoto,A., and
Sano,M., Miyata,A., Mitani,Y., Iida,K., Uesugi,H., Sugiyama,Y. and
Nomoto,H.
TITLE Expressed genes in C.elegans
JOURNAL Unpublished (1999)
COMMENT On May 9, 1996 this sequence version replaced gi:1133045.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/map="between D11S1765 and UGB"

FEATURES
source
Email: ykohara@lab.nig.ac.jp.
Location/Qualifiers
1. .300
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/map="between D11S1765 and UGB"

```

```

/clone="yk681h6"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/dev_stage="embryo"
BASE COUNT 75 a 68 c 77 g 75 t 5 others
ORIGIN

alignment_scores:
Quality: 66.50 Length: 42
Ratio: 2.293 Gaps: 1
Percent Similarity: 69.048 Percent Identity: 35.714

alignment_block:
US-09-209-961-7 x AV185677/rev ..
Align seg 1/1 to reverse of: AV185677 from: 1 to: 300
1 MetLysCysThrLeuValPheGlnSerArgPhe...CysIlePheProle 16
||||| ||||| ::||| ||| |||||::|||
204 ATGGCTTCAACTCTGAGTATCCAGTGGAGAACTGATCGGTGATGCCAAG 155
16 uThrPheLysSerAlaSerProArgLysPheLeuThrAsnValThrG 33
:::::|||||: ::||| |||||::|
154 ATCTATCATGCTACGAGGAGACCTCGCAATTCAGCGCATGTCATCTC 105
33 lyCysCysPheAlaThrValThrArg 41
:::::|||||: ::||| |||||::|
104 GCGCATGCTCTCGGACATTCGCGC 79

seq_name: gb_est5:D72647

seq_documentation_block:
LOCUS D72647 323 bp mRNA EST 13-DEC-1995
DEFINITION CELK106CYR Yuji Kohara unpublished cDNA:Strain N2 hermaphrodite
embryo Caenorhabditis elegans cDNA clone yk106c11 3', mRNA
sequence.
ACCESSION D72647
NID g1112362
VERSION D72647.1 GI:1112362
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
AUTHORS Kohara,Y., Mitsuki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and
Tabara,H.
TITLE Toward an expression map of the C.elegans genome
JOURNAL Unpublished (1994)
COMMENT On May 8, 1995 this sequence version replaced gi:801481.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 81-559-81-6854
Fax: 81-559-81-6855
Email: ykohara@lab.nig.ac.jp
Insert Length: 1990 Std Error: 0.00
High quality sequence stop: 307.
Location/Qualifiers
1. .323
/organism="Caenorhabditis elegans"
/strain="N2"
/db_xref="taxon:6239"
/map="4"
/clone="yk106c11"
/clone_lib="Yuji Kohara unpublished cDNA:Strain N2
hermaphrodite embryo"
/dev_stage="embryo"

```

QM of: US-09-209-961-9 to: GenEmbl.* out_format : pfs

Date: Dec 27, 1999 2:13 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-1998 CompuGen Ltd.

Command line parameters:

```
-MODEL=frame+ p2n.model -DEV=xlp  
-Q=Cgcn2.1/USPTO.spool/US09209961/runat_22121999_101043_17910/app_query.fasta.1  
-DB=GenEmbl -QPMF=fastap -SUFFIX=ige -CAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPEXT=0.000 -GAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human4.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=score -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pfs -NORM=stat -USER=US09209961 -NCPU=6 -ICPU=3 -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

Query: US-09-209-961-9
Query length: 104
Database: GenEmbl.*
Database sequences: 780561
Database length: 2137953050
Search time (sec): 1994.760000

score_list:

Sequence	strd	Orig	zscore	EScore	Len	Documentation
gb_v1:AF085695	-	572.00	1150.26	4.5e-56	1768	AF085695 Porcine circovirus st
gb_v1:AF086834	-	572.00	1150.26	4.5e-56	1768	AF086834 Porcine circovirus st
gb_v1:AF086835	-	572.00	1150.26	4.5e-56	1768	AF086835 Porcine circovirus st
gb_v1:AF086836	-	572.00	1150.26	4.5e-56	1768	AF086836 Porcine circovirus st
gb_v1:AF118095	-	541.00	1096.59	4.4e-53	597	AF118095 Porcine circovirus ty
gb_v1:AF027217	-	541.00	1086.76	1.6e-52	1768	AF027217 Porcine circovirus st
gb_v1:AF055391	-	541.00	1086.76	1.6e-52	1768	AF055391 Porcine circovirus ty
gb_v1:AF055392	-	541.00	1086.76	1.6e-52	1768	AF055392 Porcine circovirus ty
gb_v1:AF109397	-	541.00	1086.76	1.6e-52	1768	AF109397 Bovine circovirus, c
gb_v1:AF147751	-	541.00	1086.76	1.6e-52	1768	AF147751 Porcine circovirus ty
gb_v1:AF109399	-	535.00	1074.47	7.6e-52	1768	AF109399 Porcine circovirus ty
gb_v1:PCAJ3185	-	532.00	1068.32	1.7e-51	1768	AJ223185 Porcine circovirus ty
gb_v1:AF112862	-	531.00	1066.27	2.2e-51	1768	AF112862 Porcine circovirus ty
gb_v1:AF154679	-	522.00	1047.84	2.3e-50	1768	AF154679 Porcine circovirus, c
gb_v1:AF166528	-	522.00	1047.84	2.3e-50	1768	AF166528 Porcine circovirus, c
gb_v1:AF055393	-	520.00	1043.74	3.9e-50	1767	AF055393 Porcine circovirus ty
gb_v1:AF055394	-	520.00	1043.74	3.9e-50	1767	AF055394 Porcine circovirus ty
gb_v1:AF117753	-	511.00	1025.30	4.1e-49	1768	AF117753 Porcine circovirus ty
gb_v1:AF109398	-	506.00	1015.06	1.5e-48	1768	AF109398 Porcine circovirus ty
gb_v1:AF071879	-	278.00	548.05	1.6e-22	1758	AF071879 Porcine circovirus, c
gb_v1:PCOMGEN	-	278.00	548.04	1.6e-22	1759	Y09921 Porcine circovirus comp
gb_v1:AF012107	-	276.00	543.94	2.7e-22	1759	AF012107 Porcine circovirus, c
gb_v1:PCU49186	-	276.00	543.94	2.7e-22	1759	U49186 Porcine circovirus, com
gb_v1:HS28C9R	-	74.00	145.60	4.14	319	Z58065 H.sapiens CpG island DNA
gb_v1:AF074399	-	75.50	141.38	7.12	714	AF074399 Medicago truncatula in
gb_v1:HSU79241	-	77.50	139.32	9.28	1410	U79241 Human clone 23759 mRNA,
gb_v1:AF055259	-	71.50	138.48	10.33	398	A005259 Rattus norvegicus, OTS
gb_v1:HUMLCT09	-	71.50	138.06	10.91	417	M61842 Human lactase phlorizin
gb_v1:AF103777	-	72.50	137.12	12.31	580	AF103777 Pepper mild mottle vir
gb_v1:AF118935	-	69.00	135.71	14.74	307	AF118935 Mus musculus AE0018 in
gb_v1:AF0235274	-	71.50	135.02	16.10	583	A0235274 Itersonilia perplexans
gb_v1:AF053170	-	72.00	133.95	18.48	735	AF053170 Trifolium pallescens l
gb_v1:MTVCP	-	69.50	132.80	21.41	474	D13367 Tobacco mosaic virus CP
gb_v1:HSAL132WE9	-	67.00	130.58	28.46	344	Z24643 H.sapiens (D16S517) DNA
gb_v1:CMU45620	-	68.50	130.53	28.66	486	U45720 Candida mesenterica 26S
gb_v1:GGA240695	-	68.50	130.29	29.56	499	AF240695 Gallus gallus DNA for
gb_v1:AF152942	-	68.50	129.74	31.70	530	AF152942 Clavospora opuntiae st
gb_v1:MTVCOTATRA	-	69.50	129.56	32.45	678	L35073 Tobacco mosaic virus coa
gb_v1:MTVCQATPRO	-	69.50	129.53	32.56	680	L35074 Tobacco mosaic virus coa
gb_v1:AB012214	-	78.50	129.45	32.89	5253	AB012214 Rattus rattus mRNA fo
gb_v1:PMNVICP	-	69.50	129.28	33.62	699	X72587 Pepper mild mottle virus
gb_v1:EL3985	-	73.50	129.08	34.49	1766	EL3985 Arabidopsis thaliana mR

gb_ov:MIEFCYTB + 71.50 128.93 35.19 1143 ! X74253 E.fastuosus mitochon
gb_pr1:D87058 + 70.00 128.75 35.99 830 ! D87058 Human DNA, endogenous
gb_ro:MMVH0001 + 66.00 128.69 36.26 338 ! X53338 Mouse mRNA for the V

seq_name: gb_v1:AF085695

seq_documentation_block:

LOCUS AF085695 1768 bp DNA circular VRL 30-SEP-1998
DEFINITION Porcine circovirus strain 412, complete genome.
ACCESSION AF085695
NID 93688362
VERSION AF085695.1 GI:3668362
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
Porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
Emergence of a new porcine circovirus
Unpublished
JOURNAL
REFERENCE
2 (bases 1 to 1768)
Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
Direct Submission
TITLE
Submitted (21-AUG-1998) Veterinary Infectious Disease Organization,
University of Saskatchewan, 120 Veterinary Road, Saskatoon,
Saskatchewan S7N 5E3, Canada
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus"
/strain="412"
/db_xref="taxon:46221"
51..995
/codon_start=1
/product="P35.8"
/protein_id="AAC61860.1"
/db_xref="PID:g3668363"
/db_xref="GI:3668363"
/translation="MPSKNGRSGPQPKRWVTLNPNSEDERKKIRELPISLDYFI
VIEGNEERTPHLQGFANFKQTFNKVKGARCHIEKAGTDOONKEICSKEN
LIEGAPNSVHVSQDSLTAVSLLEGILTVAKQHPVTFVFNFRGLAEILRVSGKM
QRDKNTYHFIVGPPSCSKSWANFANPETTYWPKPKMWDGYGHEKVVVDDFY
GWLPMDDLRLCDRYPLTVKTKGTVFPLARSILITSNQPLEWYSTAVPAVEALYR
RITSLVFNKATKQSTEEGQFVTLSPPCPEFFEYNY"
complement(357..671)
/codon_start=1
/product="P12.1"
/protein_id="AAC61865.1"
/db_xref="PID:g3668368"
/db_xref="GI:3668368"
/translation="MTIIPLVRFVPCGFRVCKISSPFAFTTPRPHNEVIGFPI
TLLHFFAHQKFSQPAIEFDKRYRLLCNGHQPALQQGTHSSRQVTPLSLRKSSTF
N"
complement(386..565)
/codon_start=1
/product="P6.6"
/protein_id="AAC61866.1"
/db_xref="PID:g3668369"
/db_xref="GI:3668369"
/translation="MKCTLVFQSRFCIFPLTFKSSASPRKFLNVTGCCFATVTRIP
L
SNKVLTAVDRLRCP"
553..732
/codon_start=1
/product="P6.6"
/protein_id="AAC61861.1"
/db_xref="PID:g3668364"
/db_xref="GI:3668364"
/translation="MTLSWGLGVVKGANGLLILQTRPKHTGNHLKTSGGWMTVKKKW
LLMLTFMACRCGMV"
P33..970
/codon_start=1
/product="P5.0"
/protein_id="AAC61862.1"
/db_xref="PID:g3668365"

FEATURES

source

CDS

CDS

CDS

CDS

CDS

/db_xref="GI:3668365"
/translation="MVLNCCPSRSLSEDFLGLILECYKTIHGGRGVPVRHPFPPM
P"

CDS

1016..1177
/codon_start=1
/product="P6.2"
/protein_id="AAC61863.1"
/db_xref="PID:g3668366"
/db_xref="GI:3668366"
/translation="MFLIHLGFKWGVKIKFSELYIHGYTDIWLVLVFTVFRSAA
YVHISGL"

CDS

complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61864.1"
/db_xref="PID:g3668367"
/db_xref="GI:3668367"
/translation="MTYPRRYRRRRHPSHLGQILRRPWLPHRRHRYWRKNGI
FNTLSRTFGYTKRTVTTPSWAVDMRKKIDDVPPGGTGNKISIPFYYIRKVK
VETWPCSPITQGRGVGSTAVILDNFTKATALTYPVYNSRHTIPOPFYSHSRY
FTPKVLDSDIDYFQPNKRNQLWLRQTSNVDHVLGTAFENSKYDQDYNIRVTMY
VQREFNLKDPPLP"

BASE COUNT
ORIGIN

463 a 362 c 481 g 462 t

alignment_scores:

Quality: 572.00 Length: 104
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-9 x AF085695/rev ..

Align seg 1/1 to reverse of: AF085695 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPheProValCysGlyPh 17

|||||
671 ATGGTAACCATCCACCACCTGTTTATAGTGGTTCCAGTATGTTT 622

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTTP 34

|||||
621 CCGGGTTTGCAAAATTAGCAGCCATTGCTTTTACCACACCCAGGTGC 572

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50

|||||
571 CCCAATGAAGTACATGTTTCCCAATCACGCTTTTGCATTTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67

|||||
521 GCTCACTTCAAAAGTTACAGCCAGCCGCGGAAATTTTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84

|||||
471 CAGGTGCTGCTTTGCAACGGTCACAGAAATCCCGCTCTCCAACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100

|||||
421 CTCACAGCAGTAGACAGGTGCTGCTTCTTGCATGAGAGGAGCTCC 372

101 ThrPheAsnLys 104

371 ACATTCAATAAG 360

seq_name: gb_v1:AF086834

seq_documentation_block: 1768 bp DNA circular VRL 29-SEP-1998

LOCUS AF086834

DEFINITION Porcine circovirus strain B9, complete genome.

ACCESSION AF086834

NID 93661515

VERSION AF086834.1 GI:3661515

KEYWORDS

SOURCE

porcine circovirus.
porcine circovirus
Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE

1 (bases 1 to 1768)
Wang, L., Willison, P., Chow, B., Gibbons, E. and Babiuk, L.
Emergence of a new porcine circovirus
Unpublished

JOURNAL

REFERENCE

2 (bases 1 to 1768)
Wang, L., Willison, P., Chow, B., Gibbons, E. and Babiuk, L.
Direct Submission
Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada

JOURNAL

FEATURES

Location/Qualifiers
1..1768
/organism="porcine circovirus"
/strain="B9"
/db_xref="taxon:46221"
51..995

CDS

/codon_start=1
/product="P35.8"
/protein_id="AAC61737.1"
/db_xref="PID:g3661516"
/db_xref="GI:3661516"
/translation="MPSKNGRSGPQPHKRVFTINNPSEDERKKIRELPISLDFYFI
VGEENEGRTPHLQGFANFVKQTFNKVKYLGAICHIEKAKGTDOQNKCYCSKEGN
LLIEGAPRSQGRSDILSTAVSTLLESGILVTAKQHPVTFVKNFRGLAELLKYSKGM
OKRDWKTNVHIVGPPCGGSKWAANFANPETTYKPKNKNWDGYHGEKVYVIDFY
GWLVDLLRLCDRIPLTVTKGTGTVPLKARSLITISNQTPLEWTSSTAVFAVEALIR
RITSLVFNKNVTEQTEEGQFVTLSPPCPEFFEYIN"
complement(1034..1735)
/codon_start=1
/product="P27.9"
/protein_id="AAC61738.1"
/db_xref="PID:g3661517"
/db_xref="GI:3661517"
/translation="MTYPRRYRRRRHPSHLGQILRRPWLPHRRHRYWRKNGI
FNTLSRTFGYTKRTVTTPSWAVDMRKNIDDVPPGGTGNKISIPFYYIRKVK
VETWPCSPITQGRGVGSTAVILDNFTKATALTYPVYNSRHTIPOPFYSHSRY
FTPKVLDSDIDYFQPNKRNQLWLRQTSNVDHVLGTAFENSKYDQDYNIRVTMY
VQREFNLKDPPLP"

CDS

BASE COUNT 463 a 357 c 482 g 466 t

ORIGIN

alignment_scores:

Quality: 572.00 Length: 104
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-9 x AF086834/rev ..

Align seg 1/1 to reverse of: AF086834 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPheProValCysGlyPh 17

|||||
671 ATGGTAACCATCCACCACCTGTTTATAGTGGTTCCAGTATGTTT 622

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTTP 34

|||||
621 CCGGGTTTGCAAAATTAGCAGCCATTGCTTTTACCACACCCAGGTGC 572

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50

|||||
571 CCCAATGAAGTACATGTTTCCCAATCACGCTTTTGCATTTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67

|||||
521 GCTCACTTCAAAAGTTACAGCCAGCCGCGGAAATTTTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84

|||||
471 CAGGGTGTGCTTTTGCAACGGTCACCAAGATCCCGCTCTCCAACAGGTA 422

```

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
421 CTCACAGCAGTAGACAGGTCACTCGGTGCTCTGAGATCGAGGAGCTCC 372
|||||
101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360
seq_name: gb_vi:AF086835

seq_documentation_block:
LOCUS AF086835 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain 9741, complete genome.
ACCESSION AF086835
NID 93661518
VERSION AF086835.1 GI:3661518
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL Unpublished
TITLE Emergence of a new porcine circovirus
REFERENCE
2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL Direct Submission
TITLE Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
JOURNAL Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus"
/strain="9741"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="p35.8"
/protein_id="AAC61739.1"
/db_xref="PID:g3661519"
/translation="MPSKNGSGPQPHKRWFTLNNPSENKRKKIRELPISLFDYFI
VEEGNEEGRTPHLOGFANFKVKTENKVLGARGHIEKAKGTDOQNKKEYCSKEGN
LLIECGAPRSQGSQSDLSAVSTLLESGLITVAKQHPVTFVKNFGLAELLKVSQGM
QKRDKNVHFIVGPPCGSKSWANFANPETTYKPPKKNWDVGHGKVVVIDFY
GWLPMDDLRLCDRIPLTVKTKGTVFPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKNATQSTEEGGQFVTLSPPCPEFFEYNY"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61740.1"
/db_xref="PID:g3661520"
/db_xref="GI:3661520"
/translation="MTYPRRRYRRRRHRRPSHLGQILRRRPLVHPRHRYRWRKNGI
FNTRLRSTGYTVKRTVTTPSWAVDMREKIDDFVPPGGTNKISIPFEYVIRKVK
VEFPCSPITQDGRGVGSFVILDDNFVKATALYDYPVYNSRHTIPQPSYHSRY
FTPKPVLDDSTIDIFQPNRNQLWLQTSNVDHVLGLTAFENSKYDQDYNIRVTMY
VQFRFNKDPPLKP"
BASE COUNT 464 a 360 c 480 g 464 t
ORIGIN

alignment_scores:
Quality: 572.00 Length: 104
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-209-961-9 x AF086835/rev ..
Align seg 1/1 to reverse of: AF086835 from: 1 to: 1768
1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyP 17

```

```

|||||
671 ATGGTAACCATCCACCACTGTTTTTAGGTGGTTTCCAGTATGTGGTTT 622
|||||
17 eargValCysLysIleSerSerProPheAlaPheThrThrProArgTrpP 34
|||||
621 CCGGGTTTGCAAAATTAGCAGCCCATTTGCTTTTACCACACCCAGGTGGC 572
|||||
34 rOHISAsnGluValTrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCACAATGAAGTGATACATGTTTCCCAATCACGCTTTTGCAATTTCCC 522
|||||
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GTCACACTTCAAAAGTTCAGCCAGCCCGGAAATTTTGCACAAACGTTA 472
|||||
67 rargValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
|||||
471 CAGGGTGCTGCTTTGCAACGGTCACCAAGATCCGCTCTCCACAAGGTA 422
|||||
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
421 CTCACAGCAGTAGACAGGTCACTCGGTGCTCTGAGATCGAGGAGCTCC 372
|||||
101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360
seq_name: gb_vi:AF086836

seq_documentation_block:
LOCUS AF086836 1768 bp DNA circular VRL 29-SEP-1998
DEFINITION Porcine circovirus strain M226, complete genome.
ACCESSION AF086836
NID 93661521
VERSION AF086836.1 GI:3661521
KEYWORDS
SOURCE
ORGANISM
porcine circovirus.
VIRUSES; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL Unpublished
TITLE Emergence of a new porcine circovirus
REFERENCE
2 (bases 1 to 1768)
AUTHORS Wang,L., Willson,P., Chow,B., Gibbons,E. and Babiuk,L.
JOURNAL Direct Submission
TITLE Submitted (26-AUG-1998) VIDO, University of Saskatchewan, 120
JOURNAL Veterinary Road, Saskatoon, Saskatchewan S7N 5E3, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus"
/strain="M226"
/db_xref="taxon:46221"
51. .995
/codon_start=1
/product="p35.8"
/protein_id="AAC61741.1"
/db_xref="PID:g3661522"
/db_xref="GI:3661522"
/translation="MPSKNGSRGPQPHKRWFTLNNPSEDERKKIRELPISLFDYFI
VEEGNEEGRTPHLOGFANFKVKTENKVLGARGHIEKAKGTDOQNKKEYCSKEGN
LLIECGAPRSQGSQSDLSAVSTLLESGLITVAKQHPVTFVKNFGLAELLKVSQGM
QKRDKNVHFIVGPPCGSKSWANFANPETTYKPPKKNWDVGHGKVVVIDFY
GWLPMDDLRLCDRIPLTVKTKGTVFPFLARSILITSNQTPLEWYSSTAVPAVEALYR
RITSLVFNKNATQSTEEGGQFVTLSPPCPEFFEYNY"
complement(1034..1735)
/codon_start=1
/product="p27.9"
/protein_id="AAC61742.1"
/db_xref="PID:g3661523"
/db_xref="GI:3661523"
/translation="MTYPRRRYRRRRHRRPSHLGQILRRRPLVHPRHRYRWRKNGI
FNTRLRSTGYTVKRTVTTPSWAVDMREKIDDFVPPGGTNKISIPFEYVIRKVK

```

VEFWPCSPITQGRGVSTAVILDNEFTKATALTYDPYVYNSRRHTIPOPFYSHSRY
FTKPLVLSIDYFQPNKRNQLWLRIQTSGNVHDVGLGTAFENSKYDQDYNIRVTWY
VFREFNKDPLEP"

BASE COUNT 463 a 363 c 480 g 462 t
ORIGIN

alignment_scores:
Quality: 572.00 Length: 104
Ratio: 5.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-209-961-9 x AF086836/rev ..

Align seg 1/1 to reverse of: AF086836 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACCTGTTTGTAGTGTTCCAGTATGTGGTTT 622

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTrpP 34
|||||
621 CCGGGTTTGC AAAATAGCAGCCCATTTGCTTTTACACACCCAGGTGC 572

34 roHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCACAAAGTAAGTGTACATGTTTTCCTTCCAAATCAGCTTTTGCAATTTCCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTCAAAGTTTACAGCCAGCCGCGGAAATTTTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
|||||
471 CAGGTTGCTGTTTGCAGCGTCACAGAAATCCGCTCTCCAAACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
421 CTCACAGCAGTAGACAGGTCTCCGTTGCTTGTAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360

seq_name: gb_vi:AF118095

seq_documentation_block:

LOCUS AF118095 597 bp DNA VRL 27-JUL-1999
DEFINITION Porcine circovirus type 2 isolate IAF-614 unknown gene.

ACCESSION AF118095

NID 95596425

VERSION AF118095.1 GI:5596425

KEYWORDS

SOURCE

porcine circovirus type 2.
porcine circovirus type 2.
Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE

1 (bases 1 to 597)

AUTHORS Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.

TITLE Multiplex PCR for detection and typing of porcine circoviruses

J. Clin. Microbiol. (1999) In press

JOURNAL

REFERENCE

2 (bases 1 to 597)

AUTHORS Ouardani,M., Wilson,L., Jette,R., Montpetit,C. and Dea,S.

TITLE Direct Submission

Submitted (06-JAN-1999) CRMB, INRS-IAF, 531, Boulevard des

JOURNAL Prairies, Laval, Que H7N 4Z3, Canada

FEATURES

Location/Qualifiers

1..597

/organism="porcine circovirus type 2"

/isolate="IAF-614"

/db_xref="taxon:85708"

/note="virus isolated from lung, lymph node, spleen and tonsil tissue from pigs affected by postweaning

CDS

multisystemic wasting syndrome"

<1..>597

/note="ORE1"

/codon_start=1

/product="unknown"

/protein_id="AAD45580.1"

/db_xref="PID:g5596426"

/db_xref="GI:5596426"

/translation="SEDERKKIRELPSISLDYFIVGEGNEGRTPHLOGFANFVKQ
TFNKVYVLGARCHEIAKAGTDHONKEYCKEKNLLIECCAPRSQQRSDLSAVSTL
LESGSLVTVAQHPVTFVFNFRGLAELLKVGSKMRKDKMTNHNHVIIVGPPGCKSKWA
ANFADPETTYKPKRNKMGVGYHGEVVVIDDFYGLPW"

BASE COUNT 169 a 112 c 170 g 146 t

ORIGIN

alignment_scores:

Quality: 541.00 Length: 104

Ratio: 5.304 Gaps: 0

Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:

US-09-209-961-9 x AF118095/rev ..

Align seg 1/1 to reverse of: AF118095 from: 1 to: 597

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
549 ATGGTAACCATCCACCACCTGTTTCTAGTGGTTTCCAGTATGTGGTTT 500

17 eArgValCysLysIleSerSerProPheAlaPheThrProArgTrpP 34
|||||
499 CCGGGTCTGCAAAATAGCAGCCCATTTGCTTTTACACACCCAGGTGC 450

34 roHisAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
449 CCACAAATGACGTGTACATTTGCTTCCAATCAGCTTCTGCATTTTCCC 400

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
399 GCTCACTTTCAAAGTTTACAGCCAGCCGCGGAAATTTTCTGACAAACGTTA 350

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
|||||
349 CAGGTGCTGCTCTGCAACGTCACACAGATCCCGCTCTCCAAACAGTA 300

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||
299 CTCACAGCAGTAGACAGGTCACTCCGTTGCTTGTAGATCGAGGAGCTCC 250

101 ThrPheAsnLys 104

|||||

249 ACATTCAATAAG 238

seq_name: gb_vi:AF027217

seq_documentation_block:

LOCUS AF027217 1768 bp DNA circular VRL 14-MAY-1998
DEFINITION Porcine circovirus strain pmws PCV, complete genome.

ACCESSION AF027217

NID 92689645

VERSION AF027217.1 GI:2689645

KEYWORDS

porcine circovirus.

porcine circovirus

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE

1 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.

TITLE Nucleotide sequence of porcine circovirus associated with

postweaning multisystemic wasting syndrome in pigs

J. Virol. 72 (6), 5262-5267 (1998)

JOURNAL 98241772

REFERENCE 2 (bases 1 to 1768)

AUTHORS Hamel,A.L., Lin,L.L. and Nayar,G.P.S.
 TITLE Direct Submission
 JOURNAL Submitted (26-SEP-1997) Virology Laboratory, Veterinary Services Branch, Manitoba Agriculture, 545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada

FEATURES
 source location/Qualifiers
 1. .1768
 /organism="porcine circovirus"
 /strain="pmws PCV"
 /db_xref="taxon:46221"
 /note="both strands of seven overlapping PCR fragments were sequenced; virus isolated from lung, lymph node, spleen and tonsil tissue from pigs affected by post weaning multisystemic wasting syndrome"
 51. .995
 /note="ORF1: similar to Rep protein encoded by non-pathogenic PCV, GenBank Accession Number U49186; predicted 35.8 kDa protein"
 /codon_start=1
 /product="putative Rep protein"
 /protein_id="AAC59462.1"
 /db_xref="PID:g2689646"
 /db_xref="GI:2689646"
 /translation="MPSKNGRSGQPQPHKRWFTLNPNSEDERKKIRELPISLDYFI VGESEGEERTPHLOGFANFVKQTQNKVYLGARCYIEKAGTDQONKEYCSKEGN LITECAPRSQSGSLSTAVSLTESGLTVAEQHPVTFFVRNFRGLAEVLKVSQGM QKRDWNTNHHVIVGPPCGCKSKWAANFADPTTYWKPNNKWDGKHGEVNVVIDIFY GWPNDLLRLCDRPLTVETKGTVPFLARSILITSNQTPLEWYSSTAVPAVEALYR RITSLVFNKATEQSTEEGGQFVTLSPCCPEFFYEINY"

CDS
 117. .125
 /note="glycosylation site"
 complement(357. .671)
 327. .332
 /note="ORF3; predicted 11.9 kDa protein"
 /codon_start=1
 /protein_id="AAC59464.1"
 /db_xref="PID:g2689648"
 /db_xref="GI:2689648"
 /translation="MTYIPLVSRFPVCGFRVCKYISSPFAFTTPRPHNDVYIGLPI TLHFFAHFKFSQPAEISDKRYRVLICNGHQTALQOQTHSSRQVTPLSRSTSF NK"

CDS
 complement(386. .565)
 /note="ORF4; predicted 6.5 kDa protein"
 /codon_start=1
 /protein_id="AAC59465.1"
 /db_xref="PID:g2689649"
 /db_xref="GI:2689649"
 /translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRLPL SNKVLTAVDRLRCP"

misc_feature
 complement(470. .478)
 /note="glycosylation site"
 complement(688. .753)
 /note="ORF8; predicted 2.3 kDa protein"
 /codon_start=1
 /protein_id="AAC59469.1"
 /db_xref="PID:g2689653"
 /db_xref="GI:2689653"
 /translation="MDIDHTVSDHPTAASHKSHQ"

misc_feature
 816. .824
 /note="glycosylation site"
 905. .914
 /note="glycosylation site"
 983. .988
 complement(989. .1033)
 /note="ORF1; predicted 1.8 kDa protein"
 /codon_start=1
 /protein_id="AAC59472.1"
 /db_xref="PID:g2689656"
 /db_xref="GI:2689656"
 /translation="MKNNHYEIKKTQ"

polyA_signal
 1016. .1177
 /note="ORF5; predicted 6.2 kDa protein"
 /codon_start=1

misc_feature
 complement(1022. .1027)
 complement(1034. .1735)
 /note="ORF2; predicted 27.8 kDa protein"
 /codon_start=1
 /protein_id="AAC59463.1"
 /db_xref="PID:g2689647"
 /db_xref="GI:2689647"
 /translation="MTYPRRYRRRRHRRSHLQILRRRPLVHPRHRYRRWRKNGI FNTLSRTFGYTKATVTPSWADMMRFNIDFVPPGGTGNKISIPFEYVIRKVK VEFWPCSPITQDGRGVGSTAVILDDNFVTKATLTDPYVNYSSRHTTPOPEFSYHSRY FTKPVLDSIDYEQPNKRTQLWLRLQTSRNVHVGLTAFENSIYQDYNIRVTMY VQREFNLKDPPLAP"

CDS
 complement(1301. .1309)
 /note="glycosylation site"
 complement(1522. .1611)
 /note="ORF6; predicted 3.1 kDa protein"
 /codon_start=1
 /protein_id="AAC59467.1"
 /db_xref="PID:g2689651"
 /db_xref="GI:2689651"
 /translation="MASSTPASPSDILSRLPQSERPPGRWT"
 1524. .1631
 /note="ORF10; predicted 4.1 kDa protein"
 /codon_start=1
 /protein_id="AAC59471.1"
 /db_xref="PID:g2689655"
 /db_xref="GI:2689655"
 /translation="MSTAQEGVLTVVTPVKVRRVRLKMPFELLOR"

CDS
 1682. .1741
 /note="ORF7; predicted 1.9 kDa protein"
 /codon_start=1
 /protein_id="AAC59468.1"
 /db_xref="PID:g2689652"
 /db_xref="GI:2689652"
 /translation="MAAGAVSSSAVTPPWIRHS"

CDS
 complement(join(1732. .1768,1. .92))
 /note="ORF9; predicted 4.6 kDa protein"
 /codon_start=1
 /protein_id="AAC59470.1"
 /db_xref="PID:g2689654"
 /db_xref="GI:2689654"
 /translation="MWLGSSASSILLAGHVAEVLPRCCRCRSALVILTAHFFRFLQ"

stem_loop
 join(1746. .1768,1. .13)
 /function="putative replication site"

rep_origin
 join(1762. .1768,1. .2)
 /note="AAGATTAC" is similar to the nonanucleotide motif in the non-pathogenic PCV, GenBank Accession Number U49186"

BASE COUNT 452 a 360 c 495 g 461 t
 ORIGIN

alignment_scores:
 Quality: 541.00 Length: 104
 Ratio: 5.304 Gaps: 0
 Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:
 US-09-209-961-9 x AF027217/rev ..
 Align seg 1/1 to reverse of: AF027217 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyPh 17
 |||||
 671 ATGGTAACCATCCACCACTGTTCTTAGGTGGTTCCAGTATGTGTTT 622
 |||||
 17 eargValCysLysIleSerProphealaPheThrProArgTrp 34
 |||||

```

621 CCGGCTGCAAAATTAGACGCCCAATTTGCTTTTACCACACCCAGGTGGC 572
34 roHisAsnGluValThrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
571 CCACAAATGACGTGATCATGGTCTTCCCAATCACGCTCTGCATTTCCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
521 GCTCACTTCAAAAGTTCACCGACCGCGGGAATTTCTGACAAAGCTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
471 CAGGTGCTGCTCTGCAACGGTCACACAGACTCCCGCTCTCCACAAAGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
421 CTCACAGCAGTACAGAGGTCACTCCGTTGCTTGAGATCGAGAGGTCC 372
101 ThrPheAsnLys 104
|||||:|||||
371 ACATTCAATAAG 360
seq_name: gb_vi:AF055391
seq_documentation_block:
LOCUS AF055391 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from USA, complete genome.
ACCESSION AF055391
NID 93598796
VERSION AF055391.1 GI:3598796
KEYWORDS porcine circovirus.
SOURCE porcine circovirus.
ORGANISM Viruses; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Meehan,B.M., McNelly,F., Todd,D., Kennedy,S., Jewhurst,V.A., Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.
TITLE Characterization of novel circovirus DNAs associated with wasting syndromes in pigs
JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)
MEDLINE 98418498
REFERENCE 2 (bases 1 to 1768)
AUTHORS Meehan,B.M.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The Queen's University of, Belfast, Stoney Road, Belfast, Northern Ireland, BT4 3SD, U.K.
FEATURES
source
1. 1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.999 ECGRI No.8"
/country="USA"
/complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35308.1"
/db_xref="PID:g3598807"
/db_xref="GI:3598807"
/translation="MASSTPASPAPSDILSLRPLQSPERPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35306.1"
/db_xref="PID:g3598805"
/db_xref="GI:3598805"
/translation="MSTAQEGVLTVVALTVPKVRERRVLKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35307.1"

```

```

/db_xref="PID:g3598806"
/db_xref="GI:3598806"
/translation="MAAGAVSSSANTPPWIRHS"
/complement(311..439)
/notes="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35304.1"
/db_xref="PID:g3598803"
/db_xref="GI:3598803"
/translation="MWLGSSSILLAGHVAEEVLPCCRCRSALVILTAAHFFRQL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35298.1"
/db_xref="PID:g3598797"
/db_xref="GI:3598797"
/translation="MPSKNGRSGPQPHKRWVFTLNNPSEDEKRIKIELPISLDFYFI
VGEENEGRTPLQGFANFVKQTENKVKVYLGAICHIEKAGTQONKEYSKEGN
LLIEGAPRSQGSRLSTAVSTLLESGSLVTVAEOHPVTFVFNFRGLAELLKVSQKM
QKDMKTNVHVIVGPPCGCKSKWAANFADPETYWKPPRNKWDGTVHGEVVDIDFY
GWLPDLDRLCDRIPLTVETKGTVPFLARSILITSNQTPLEWISSTAVPAVEALYR
RITSLVFWKNATEQSTEGGFVTLSPCCPEPFYEINY"
674..679
/complement(704..1018)
/notes="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35300.1"
/db_xref="PID:g3598799"
/db_xref="GI:3598799"
/translation="WVTIPLVSRWFPVCGFRVCKISSPFAFTTPRPHNDVYIGLPI
TILHFPAPHFQKFSQPAEISDKRYRVLCLNGHQTALQOQTHSSRQVTPLSLRSSSTF
NK"
/complement(733..912)
/notes="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35301.1"
/db_xref="PID:g3598800"
/db_xref="GI:3598800"
/translation="MTCITLVFSQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRPL
SNKVLTAVDRLRCP"
/complement(1035..1100)
/notes="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35303.1"
/db_xref="PID:g3598802"
/db_xref="GI:3598802"
/translation="MDIDHTVSDHPTAASHKSHQ"
1330..1335
/complement(1336..1380)
/notes="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35305.1"
/db_xref="PID:g3598804"
/db_xref="GI:3598804"
/translation="MNNKNHYEVIKKTQ"
1363..1524
/notes="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35302.1"
/db_xref="PID:g3598801"
/db_xref="GI:3598801"
/translation="MVFIIHLGFKWGVKIKFSELVHIGTIDIVLVVYTVFERSAEA
YVWHISRGL"
/complement(1369..1374)
/complement(join(1381..1768,1..314))
/notes="predicted 27.8 kDa protein"
/codon_start=1

```

polyA_site
CDS


```
/product="ORF2"  
/protein_id="AAC35299.1"  
/db_xref="PID:g3598798"  
/db_xref="GI:3598798"  
/translation="MTYPRRYRRRRHPRSHLQILRRPWLHVHPRHYRWRKNGI  
FNRLSRTGYTKATVTPSWAVDMRFNIDDFVPPGGTNKISIPFIYRIKKVK  
VFEPSPITQDGRGVSTAVILDDNFVKATLDYDXYVYSSRHTIPQFYSYHRY  
FTPKPVLDTIDYFOPNNRNLWLRLQTSRNVDRHVLGLTAFENSIYDODYNIRVTMY  
VQFRENLKDPKLP"
```

BASE COUNT 453 a 360 c 494 g 461 t
ORIGIN

alignment_scores:

Quality: 541.00 Length: 104
Ratio: 5.304 Gaps: 0
Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:

US-09-209-961-9 x AF055391/rev ..

Align seg 1/1 to reverse of: AF055391 from: 1 to: 1768

```
1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyPh 17  
|||||  
1018 ATGCTAACCATCCACCACTGTTTCTAGTGGTTTCCAGTATGTGGTTT 969  
17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34  
|||||  
968 CCGGCTGTCGAAAATTAGCAGCCATTGCTTTTACCACACCCAGGTGCG 919  
34 roHisAsnGluValIleGlyPheProIleThrLeuLeuHisPhePro 50  
|||||  
918 CCCACATGAGCTGTACATGCTTCCATCAGCTTCTGCATTTTCCC 869  
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67  
|||||  
868 GCTCATTTCAAAAGTTTCAGCAGCCGCGGAAATTTCTGACAAAGGTA 819  
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84  
|||||  
818 CAGGCTGCTGCTGCAACGGTCCACGACTCCCGCTCTCCAAACAGGTA 769  
84 rHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100  
|||||  
768 CTCACAGCAGTAGACAGTCACTCCGTTGCTTGTGATCGAGGAGCTCC 719  
101 ThrPheAsnLys 104  
|||||  
718 ACATTCATAAG 707
```

seq_name: gb_vi.AF055392

seq_documentation_block:

LOCUS AF055392 1768 bp DNA circular VRL 13-SEP-1998
DEFINITION Porcine circovirus Type II from Canada, complete genome.

ACCESSION AF055392

NID 93598808

VERSION AF055392.1 GI:3598808

KEYWORDS

SOURCE porcine circovirus.

ORGANISM

Viruses; ssDNA viruses; Circoviridae; Circovirus.

REFERENCE 1 (bases 1 to 1768)

AUTHORS Meehan,B.M., McNeill,F., Todd,D., Kennedy,S., Jewhurst,V.A.,
Ellis,J.A., Hassard,L.E., Clark,E.G., Haines,D.M. and Allan,G.M.

TITLE Characterization of novel circovirus DNAs associated with wasting

syndromes in pigs

JOURNAL J. Gen. Virol. 79 (Pt 9), 2171-2179 (1998)

MEDLINE 98418498

REFERENCE 2 (bases 1 to 1768)

AUTHORS Meehan,B.M.

TITLE Direct Submission

JOURNAL

Submitted (26-MAR-1998) Brian M Meehan, Veterinary Science, The
Queen's University of, Belfast, Stoney Road, Belfast, Northern
Ireland, BT4 3SD, U.K.

FEATURES

source

Location/Qualifiers

1..1768
/organism="porcine circovirus"
/specific_host="Sus scrofa"
/strain="Porcine circovirus Type II"
/db_xref="taxon:46221"
/clone="Imp.1010-Stoon EcoRI No.14"
/country="Canada"
complement(101..190)
/codon_start=1
/product="ORF6"
/protein_id="AAC35319.1"
/db_xref="PID:g3598819"
/db_xref="GI:3598819"
/translation="MASSTPASPPASDILSSVPSQSRPPGRWT"
103..210
/codon_start=1
/product="ORF10"
/protein_id="AAC35317.1"
/db_xref="PID:g3598817"
/db_xref="GI:3598817"
/translation="MSTAQEGVTVVRLTVYKVRERVLKMPFFLLQR"
261..320
/codon_start=1
/product="ORF7"
/protein_id="AAC35318.1"
/db_xref="PID:g3598818"
/db_xref="GI:3598818"
/translation="MAAGAVSSSAVTPPWIRHS"
complement(311..439)
/note="predicted 4.6 kDa protein"
/codon_start=1
/product="ORF9"
/protein_id="AAC35315.1"
/db_xref="PID:g3598815"
/db_xref="GI:3598815"
/translation="MWLGSASSILLAGHVAEVLPRCCRCRSALVILTAHFRRFOL"
398..1342
/codon_start=1
/product="putative Rep protein (ORF1)"
/protein_id="AAC35309.1"
/db_xref="PID:g3598809"
/db_xref="GI:3598809"

CDS

polyA_signal

CDS

complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35311.1"
/db_xref="PID:g3598811"
/db_xref="GI:3598811"
/translation="MVIPLVSRWFVCGFRCKISSPFAFTTPRPHNDVYICLPI
TLHFFAHPFQFSQPAISDKRYKVLCLNGHTPALQGGTHSSRQVTPLSRSSTF
NK"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35312.1"
/db_xref="PID:g3598812"
/db_xref="GI:3598812"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRPL
SNKVIATAVDRSLRCP"
complement(1035..1100)

polyA_signal

CDS

complement(704..1018)
/note="predicted 11.9 kDa protein"
/codon_start=1
/product="ORF3"
/protein_id="AAC35311.1"
/db_xref="PID:g3598811"
/db_xref="GI:3598811"
/translation="MVIPLVSRWFVCGFRCKISSPFAFTTPRPHNDVYICLPI
TLHFFAHPFQFSQPAISDKRYKVLCLNGHTPALQGGTHSSRQVTPLSRSSTF
NK"
complement(733..912)
/note="predicted 6.5 kDa protein"
/codon_start=1
/product="ORF4"
/protein_id="AAC35312.1"
/db_xref="PID:g3598812"
/db_xref="GI:3598812"
/translation="MTCTLVFQSRFCIFPLTFKSSASPRKFLTNVTGCCSATVTRPL
SNKVIATAVDRSLRCP"
complement(1035..1100)

```

/note="predicted 2.3 kDa protein"
/codon_start=1
/product="ORF8"
/protein_id="AAC35314.1"
/db_xref="PID:93598814"
/db_xref="GI:3598814"
/translation="MDIDHTVSDPHPTAAASHKSHQ"
1330..1335
polyA_signal
CDS
complement(1336..1380)
/note="predicted 1.8 kDa protein"
/codon_start=1
/product="ORF11"
/protein_id="AAC35316.1"
/db_xref="PID:93598816"
/db_xref="GI:3598816"
/translation="MNNKHVEVIKIQ"
1363..1524
CDS
/note="predicted 6.2 kDa protein"
/codon_start=1
/product="ORF5"
/protein_id="AAC35313.1"
/db_xref="PID:93598813"
/db_xref="GI:3598813"
/translation="MVFIIHLGKMGVFKIKFSELYIHGYTDIVLVVVTTFERSAEA
YVHISRGL"
complement(1369..1374)
/note="predicted 27.8 kDa protein"
/codon_start=1
/product="ORF2"
/protein_id="AAC35310.1"
/db_xref="PID:93598810"
/db_xref="GI:3598810"
/translation="MTIPRRYRRRRHPRSHLQGLRRRLVHPHRYWRKNGI
FNTLSRTFTGTRGTVTTTPSWAVDMRFXIDDFVPGGNTKISIPFYIRKRV
VEFWPCSPITOGDGVGSTAVLDDNFVTKATLTDYDVNYSRHTIPQPFYSHSRY
FTPKPVLIDSTIDYQPNKRNQLRLQTSNGVDHVLGAAFENSKYDQYNIRVTMY
VORFEFLKDPPLAP"
BASE COUNT 451 a 361 c 495 g 461 t
ORIGIN

alignment_scores:
Quality: 541.00 Length: 104
Ratio: 5.304 Gaps: 0
Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:
US-09-209-961-9 x AF055392/rev ...
Align seg 1/1 to reverse of: AF055392 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
1018 ATGGTAACCATCCACCACCTGTTTCTAGGTGCTTCCAGTAGTGTTT 969
|||||
17 earGValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
|||||
968 CCGGGTCTGCAAAATAGCAGCCATTGCTTTTACCACACCCAGGTGC 919
|||||
34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
918 CCACAATGACGGTGCATTGGTCTTCCCAATCAGCTTCGCAATTTC 869
|||||
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
868 GCTCACTTTCARAAGTTCACGACGCCGCGGAATTCGACAAACGTTA 819
|||||
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
818 CAGGTGCTGCTGTGCAACGGTCACCACTCCCGCTCTCCAAACAGGTA 769
|||||
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100

```

```

|||||
768 CTCACAGCAGTAGACAGGTCACTCCGTTGTCTTGAGATCGAGAGCTCC 719
|||||
101 ThrPheAsnLys 104
|||||
718 ACATTCAATAAG 707
|||||
seq_name: gb_v1.AF109397
seq_documentation_block:
LOCUS AF109397 1768 bp DNA circular VRL 02-FEB-1999
DEFINITION Bovine circovirus, complete genome.
ACCESSION AF109397
NID 94210313
VERSION AF109397.1 GI:4210313
KEYWORDS
SOURCE
ORGANISM
bovine circovirus.
Virus; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE
1 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Nucleotide sequence of a circovirus detected in cattle with various
clinical syndromes
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REFERENCE
3 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (02-FEB-1999) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
REMARK Sequence update by submitter
COMMENT On Feb 2, 1999 this sequence version replaced gi:4028609.
FEATURES
location/Qualifiers
1..1768
/organism="bovine circovirus"
/virus
/db_xref="taxon:85542"
/note="BCV"
sequence was obtained from several overlapping PCRs using
DNA extracted from various tissues from cattle; similar to
Porcine circovirus sequence presented in GenBank Accession
Number AF027217"
13..35
repeat_region
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
/note="ORF-1; similar to Porcine circovirus ORF1 encoded
by the sequence presented in GenBank Accession Number
AF027217."
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD11928.1"
/db_xref="PID:94210314"
/db_xref="GI:4210314"
/translation="MPSKNGRSGPQPKRWVTLNPNSEDERKKIRLPLISLDFYFI
VGEENEGRTPHLQGFANFYKQTFNKVGLGARCHIEKAKGTQQKNEVCSEKGN
LLIEGAPRSQGRSLDSTAVSLTLESGLVTAEOHPVTFVNFNGLAEELKVSQKM
QKRWKTNVIVPGCGCKSKAAAFADPTTYKPKRNKWDGTVHGEVVIDDFY
GWLWDDLLRLVDYPLVETRGTVVPLFARSLITSLITSNOTPLEWYSTTAVPAVEALYR
RITSLVFWKNATEQSTEEGGQFVTLSPPCPEPEYIN"
327..332
complement(357..671)
/note="similar to Porcine circovirus ORF-3 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-3"
/protein_id="AAD11930.1"
polyA_signal
CDS

```

```

/db_xref="PID:g4210315"
/db_xref="GI:4210315"
/translation="MTIPLVSRWFPVCGFRVCKISSPPAFTRPWHNDVYIGLPI
TLIHPFAHQKFSQPAEISDKRYRVLLCNCHQTPALQQGTHSSRQVTPLSLRSSSTF
NK"
CDS
complement(386..565)
/Note="similar to Porcine circovirus ORF-4 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-4"
/protein_id="AADI1931.1"
/db_xref="PID:g4210316"
/db_xref="GI:4210316"
/translation="MTCTLVFQSRFCIFPLTEKSSAPRKFLNVTGCCSATVTRLPL
SNKLVAVDRSLRCP"
CDS
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AADI1927.1"
/db_xref="PID:g4028610"
/db_xref="GI:4028610"
/translation="MTSLWGLGVVYKANGLLILQTRKKPHGTGNHLETSGCMVTVKXW
LLMTFMACRGMIY"
CDS
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AADI1935.1"
/db_xref="PID:g4028618"
/db_xref="GI:4028618"
/translation="MDIDHTSVSDHPTAAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AADI1937.1"
/db_xref="PID:g4028620"
/db_xref="GI:4028620"
/translation="MNKNHYEVIKKTQ"
1016..1177
/codon_start=1
/product="ORF-5"
/protein_id="AADI1932.1"
/db_xref="PID:g4028615"
/db_xref="GI:4028615"
/translation="WVFIHLGFKWGVFKIFKSELYIHGYTDIVLVVTFVRSAAE
YVHISRGL"
polyA_signal
complement(1022..1027)
complement(1256..1735)
/Note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-2"
/protein_id="AADI1929.1"
/db_xref="PID:g4028612"
/db_xref="GI:4028612"
/translation="MTYPRRYRRTRPRSHLQILRRPWLVRPHRYWRKNGI
ENRUSRTGYTKVATTPTSWVDMRFNIDDFVPPGGTNNKISPEYIRKVK
VFVPCSPITQDGGVGYSAVILDDNFVTKATALTYDPYVYSSRHTIPQPFYSYHR"
complement(1522..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AADI1933.1"
/db_xref="PID:g4028616"
/db_xref="GI:4028616"
/translation="MASSTPASPASDILSRLPQSRPPGRWT"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AADI1936.1"
/db_xref="PID:g4028619"
/db_xref="GI:4028619"
/translation="MSTAQEGVLTVVATVYKVRERRVLKMPFFLLQR"
1682..1741
CDS

```

```

/codon_start=1
/product="ORF-7"
/protein_id="AADI1934.1"
/db_xref="PID:g4028617"
/db_xref="GI:4028617"
/translation="MAAGAGPSSAVTPPWIRHS"
521n(1750..1768,1..13)
join(1762..1768,1..2)
/Note="putative; similar to the nonnucleotide motif of
porcine circovirus presented in GenBank Accession Numbers
AF027217 and U49186"
BASE COUNT 452 a 362 c 493 g 461 t
ORIGIN

alignment_scores:
  Quality: 541.00      Length: 104
  Ratio: 5.304        Gaps: 0
  Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:
US-09-209-961-9 x AF109397/rev ..
Align seg 1/1 to reverse of: AF109397 from: 1 to: 1768

1 MetValThrIleProLeuValPheArgTTPheProValCysGlyPh 17
|||||
671 ATGGTAACATCCACCACTTGTTCAGTGGTTTCCAGATGTGGTTT 622

17 eaRgValCysLysIleSerSerProPheAlaPheThrThrProArgTTP 34
|||||
621 CGGGTCTGCAAAATAGCAGCCCATTTGCTTTACACACCAGGTGGC 572

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuHisPhePro 50
|||||
571 CCCACAATGACGTGTACATTGGTCTTCCCAATCAGGCTTCTGCATTTCC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GTCCTACTTCAAAAGTTACGCCAGCCCGCGAAATTTCTGACAAACGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGTGTGCTCTGCACAGGTCACCACTCCCGCTCTCCACACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTCACTCCGTTGCTCTTGAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACATTCATAAG 360

seq_name: gb_v1:AF147751

seq_documentation_block:
LOCUS AF147751 1768 bp DNA 08-JUN-1999
DEFINITION Porcine ci. virus type 2 unknown genes.
ACCESSION AF147751
NID 95007010
VERSION AF147751.1 GI:5007010
KEYWORDS
SOURCE
ORGANISM
REFERENCE
1 (bases 1 to 1768)
AUTHORS
Pogranichnyy, R., Yoon, K.-J., Harms, P., Swenson, S., Zimmerman, J. and
Sorden, S.
TITLE
Characterization of clinical and immune responses in young swine to
experimental porcine circovirus type II infection
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 1768)
AUTHORS
Yoon, K.-J. and Pogranichnyy, R.M.

```

TITLE Direct Submission
JOURNAL Submitted (30-APR-1999) Veterinary Diagnostic Laboratory, Iowa
State University, College of Veterinary Medicine, Ames, Iowa 50011,
USA

FEATURES

Location/Qualifiers

1. .1768
/organism="porcine circovirus type 2"
/isolate="ISUVDL 98-15237"
/specific_host="Sus scrofa"
/db_xref="taxon:85708"
/note="isolated from pigs with postweaning multisystemic
wasting syndrome"
51. .995
/note="ORF1; 35-36kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37776.1"
/db_xref="PID:g5007011"
/db_xref="GI:5007011"
/translation="MPSKNGRSGPQHKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKVKTQFNKVTGLGARCHIERAKGTDQONKEVCSEKGN
LLIECGAPRSGQSDLSSTAVSLLESGLVTAHQHPVTFVRNFRGLAELLKVSCKM
QKRDWNTVHVIVGPPGCKSKWAANFADPETTYWKPKNKWDGTHGEEVVVDIFY
GWLPMDDLRLCDRYPLTVETGKTPFLARSILITSNOTPLEWTSSTAVPAVEALYR
RITSLVFNKATEQSTEGGQFVTLSPPCPEFFPEYIN"
complement(1034. 1735)
/note="ORF2; 27-28kD"
/codon_start=1
/product="unknown"
/protein_id="AAD37777.1"
/db_xref="PID:g5007012"
/db_xref="GI:5007012"
/translation="MTYPRRYRRRRHRRSHLGOILRRRWLVHPHRYWRKNGI
ENTLSRFGTVKATVTFPSWAVDMRFNIDDFVPGCGTKNKISIFPEYIRKVK
VEFWPCSDITGDGKGVSTALNDNFVTAKATLDYDVNTSSRHTIPQPFCHSRV
FTPKPVLDSITDYFQPNKRNQLRLQTSRNVHVLGTAFENSIIYQDYNIRVTY
VQFREFNLKDPLPK"

CDS

421 CTCACAGCAGTAGACAGGTCACTCCGTTGTCTTGAGATCAGGAGCTCC 372
101 ThrPheAsnLys 104
|||||
371 ACATCAATAAG 360
seq_name: gb_v1:AF109399
seq_documentation_block:
LOCUS AF109399 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-E, complete genome.
ACCESSION AF109399
NID g4106896
VERSION AF109399.1 GI:4106896
KEYWORDS
SOURCE porcine circovirus type 2-E.
ORGANISM porcine circovirus type 2-E
Viruses; ssDNA viruses; Circoviridae; Circovirus; #69#ine
circovirus type 2.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Hamel, A.L. and Nayar, G.P.S.
TITLE Nucleotide sequence of four different isolates of circovirus
detected in pigs with various clinical syndromes
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel, A.L. and Nayar, G.P.S.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) Virology Laboratory, Veterinary Diagnostic
Laboratory, Manitoba Agriculture, 545 University Crescent,
Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
source
1. .1768
/organism="porcine circovirus type 2-E"
/virion
/db_xref="taxon:85544"
/note="similar to Porcine circovirus sequence presented in
GenBank Accession Number AF027217; type-E designation is
based upon restriction endonuclease digestion pattern;
sequence obtained from several overlapping PCRs using DNA
extracted from lung, mesenteric lymph node and tonsil of
pig"
repeat_region 13. .35
/rpt_type=tandem
/rpt_unit=13. .18
CDS 51. .995
/note="ORF-1"
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03071.1"
/db_xref="PID:g4106897"
/db_xref="GI:4106897"
/translation="MPSKNGRSGPQHKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTPLHGFANFKVKTQFNKVTGLGARCHIERAKGTDQONKEVCSEKGN
LLIECGAPRSGQSDLSSTAVSLLESGLVTAHQHPVTFVRNFRGLAELLKVSCKM
QKRDWNTVHVIVGPPGCKSKWAANFADPETTYWKPKNKWDGTHGEEVVVDIFY
GWLPMDDLRLCDRYPLTVETGKTPFLARSILITSNOTPLEWTSSTAVPAVEALYR
RITSLVFNKATEQSTEGGQFVTLSPPCPEFFPEYIN"
327. .332
complement(357. .671)
/codon_start=1
/product="ORF-3"
/protein_id="AAD03073.1"
/db_xref="PID:g4106899"
/db_xref="GI:4106899"
/translation="MTYIPPLVSRWFPVCGFRVCKISSPFAETTPRPHNDVYIGLPI
TLHFPAHFQKFSQPAEISDKRYRVLNCGHOTPALQQGTHSSRQVTPLSLRSSSTL
NK"
complement(386. .565)
/codon_start=1
/product="ORF-4"
/protein_id="AAD03074.1"
/db_xref="PID:g4106900"
/db_xref="GI:4106900"

alignment_scores:

Quality: 541.00 Length: 104
Ratio: 5.304 Gaps: 0
Percent Similarity: 98.077 Percent Identity: 95.192

alignment_block:

US-09-209-961-9 x AF147751/rev ..

Align seg 1/1 to reverse of: AF147751 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
|||||
671 ATGTAACCATCCACACACATGTTCTTAGGTGGTTTCCAGTATGTGTTT 622
17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
|||||
621 CCGGGCTGCAAAATAGACGCCATTTGCTTTACCAACCCAGGAGGC 572
34 roHisAsnGluValTrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCACAAATGACGTTGATATGGTCTTCCATCATCAGCTTCTGCATTTTCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTCAAAGTTCAGCACGCCGCGGAAATTTCTGACAAACGTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnClyT 84
|||||
471 CAGGGTGCTGCTCTGCAACGGTCACCAAGACTCCGCTCTCCAAACAGGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
|||||

```

/translation="MTCTLVQSRFCIFPLTFKSSASPRKFLTNVTVGCCSATVTRLPL
SNKLVAVDRSLRCP"
553..732
/codon_start=1
/product="ORF-12"
/protein_id="AAD03081.1"
/db_xref="PID:g4106907"
/db_xref="GI:4106907"
/translation="MTSLWGLGVKANGLLILQTRKPHNGHLETSGGMVTVKKW
LLMTFMAGCGMIV"
complement(688..753)
/codon_start=1
/product="ORF-8"
/protein_id="AAD03078.1"
/db_xref="PID:g4106904"
/db_xref="GI:4106904"
/translation="MDIDHVSVDHPTAASHKSHQ"
983..988
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03080.1"
/db_xref="PID:g4106906"
/db_xref="GI:4106906"
/translation="MNNKNHYEVIKKTQ"
1016..1363
/notes="similar to Porcine circovirus ORF-5 encoded by the
sequence presented in GenBank Accession Number AF027217."
/codon_start=1
/product="ORF-5"
/protein_id="AAD03075.1"
/db_xref="PID:g4106901"
/db_xref="GI:4106901"
/translation="MFIHLGFKWGVKIKFSELYIHGYTDIVLVTVTFERSAPA
YVHISGLSHLPQIPFVILWLEVINSIGIKRFCEVPVGVGELGDCMAGGVVYGV
IGIGCGLYKVVV"
complement(1022..1027)
complement(1034..1735)
/codon_start=1
/product="ORF-2"
/protein_id="AAD03072.1"
/db_xref="PID:g4106898"
/db_xref="GI:4106898"
/translation="MYPREFRRRRHPSHLGQILRRPWLVRPVRHYRWKKNCI
FNARLRTGYTKATVTPSPNAWMLRENLDQFVPPGGTKNISIPFIYRIRKVK
VEWPCSPITQDGRGVSSAILDDNFVIRKATQYDPYVNSRHTIPQPFYHSRY
FTPKPVLDSITIDYFQPNKRNQLMRLQTSRNDVHVLGTAFAENSKYDQDYNIRVTY
VQREFNLKDPPLKP"
1524..1631
/codon_start=1
/product="ORF-10"
/protein_id="AAD03079.1"
/db_xref="PID:g4106905"
/db_xref="GI:4106905"
/translation="MSTAQEGVLTVVATVYKVRERRALKMPFFLQFQ"
complement(1528..1611)
/codon_start=1
/product="ORF-6"
/protein_id="AAD03076.1"
/db_xref="PID:g4106902"
/db_xref="GI:4106902"
/translation="MASSTPASPAPSDILSRPQSARPPQR"
1682..1741
/codon_start=1
/product="ORF-7"
/protein_id="AAD03077.1"
/db_xref="PID:g4106903"
/db_xref="GI:4106903"
/translation="MAAGVSSSAETPPPIWRHS"
join(1750..1768,1..13)
join(1762..1768,1..2)
/notes="putative; similar to the nonanucleotide motif of
Porcine circovirus presented in Genbank Accession Numbers

```

```

BASE COUNT      448 a      362 c      495 g      463 t
ORIGIN

alignment_scores:
  Quality: 535.00      Length: 104
  Ratio: 5.245      Gaps: 0
  Percent Similarity: 98.077      Percent Identity: 94.231

alignment_block:
US-09-209-961-9 x AF109399/rev ..

Align seg 1/1 to reverse of: AF109399 from: 1 to: 1768

1 MetValThrIleProProLeuValPheArgTTPpPheProValCysGlyPh 17
|||||
671 ATGGTAACCATCCACCACTTGTCTTAGTGTTTCCAGTATGTGTTT 622

17 eargValCysLysIleSerSerProPheAlaPheThrThrProArgTripp 34
|||||
621 CCGGGTCTGCAAAATAGCAGCCCATTTGCTTTTACACACCAGGTGGC 572

34 roHisAsnGluValTyrIleGlyPheProIleThrLeuLeuHisPhePro 50
|||||
571 CCCACAATCAGCTGACATTTGGTCTTCCCAATCAGCTTCTGCATTTC 522

51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
|||||
521 GCTCACTTTTCAAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAAGTTA 472

67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGGTGCCTGCTGCAACGGTCACAGACTCCCGCTCTCCAAACAGGTA 422

84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTCACCTCCGTTGCTTGAGATCGAGGAGCTCC 372

101 ThrPheAsnLys 104
|||||
371 ACACCTCAATAAG 360

seq_name: gb_vi:PCAJ3185

seq_documentation_block:
LOCUS      PCAJ3185      1768 bp      DNA      circular      VRL      03-JUL-1998
DEFINITION Porcine circovirus DNA, complete genome, isolate ISU-31.
ACCESSION  AJ223185
NID        93293040
VERSION    AJ223185.1 GI:3293040
KEYWORDS   complete genome.
SOURCE     porcine circovirus.
ORGANISM   porcine circovirus
REFERENCE  1 (bases 1 to 1768)
AUTHORS    Morozov, I.
TITLE      Direct Submission
JOURNAL
Institute, Iowa State University, 1802 Elwood Dr. VMRI Ames, Iowa
50011, U.S.A
REFERENCE  2 (bases 1 to 1768)
AUTHORS    Morozov, I., Girinarumitr, T., Sorden, S.D., Halbur, P.G., Morgan, M.K.,
Yoon, K.J. and Paul, P.S.
TITLE      Detection of a novel strain of Porcine circovirus in pigs with
Postweaning Multisystemic Wasting Syndrome (PMWS)
JOURNAL    Unpublished
FEATURES   Location/Qualifiers
            source
            1..1768
            /organism="porcine circovirus"
            /virus
            /isolate="ISU-31"
            /specific_host="Sus scrofa"

```

```

CDS
/db_xref="taxon:46221"
51..995
/contig="ORF1"
/codon_start=1
/protein_id="CA11157.1"
/db_xref="PID:g1310034"
/db_xref="PID:g13293041"
/db_xref="GI:3293041"
/translation="MPSKNGSGPQPHKRWVTLNPSDEKRRKIRELPISLFDYFI
VGEENEGRTPHLOGFANFKVQKOTENKVKYLGARHIEKAKGTDOONKEYCSKEGN
LLIEGAPRSQGRSDLSSTAVSLLESGLVVAEOHPVTFVNRFLGAEELKVSCKM
QRDKNTNHHVIVGPPGCKSKAANFADPTTYWKPNNKWDGHEEVVDDFY
GWLWDLLRLCDRPLATYETGKTVFPLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPCPPEPYEINY"
BASE COUNT 452 a 361 c 492 g 463 t
ORIGIN

alignment_scores
Quality: 532.00 Length: 104
Ratio: 5.267 Gaps: 0
Percent Similarity: 97.115 Percent Identity: 94.231

alignment_block
US-09-209-961-9 x PCAJ3185/rev ..
Align seg 1/1 to reverse of: PCAJ3185 from: 1 to: 1768
1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyPh 17
671 ATGGTACCAATCCACACACATGTTTCTAGGTGGTTTCCAGTATGTGTTT 622
17 eArgValCysIleSerSerProPheAlaPheThrThrProArgTrpP 34
621 CCGGCTGCAAAATAGACCCATTGCTTTTACCACACCCAGGTCG 572
34 roHisAsnGlnValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
571 CCCACATGACGTGTACATTGGTCTTCCATCAGCTTCGATTTCCTCC 522
51 AlAHisPheGlnIlyPheSerGlnProAlaGluIlePheAspIlySer 67
521 GCTCACATTCACAAAGTTTCAGTACGCGCGGAAATTTCTGACAAACG 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGly 84
471 CAGGTCGCTCTGCAACGCTCACCAGACTCCGCGCTCTCCAAACAGGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
421 CTCACACAGTACAGAGTCACTCCGTTGCTTGTGATCGAGAGCTCC 372
101 ThrPheAsnIly 104
371 ACATTCAATAAG 360
seq_name: gb_v1:AF112862
seq_documentation_block:
LOCUS AF112862 1768 bp DNA circular VRL 06-JAN-1999
DEFINITION Porcine circovirus type 2-B, complete genome.
ACCESSION AF112862
NID 94106915
VERSION AF112862.1 GI:4106915
KEYWORDS
SOURCE
ORGANISM porcine circovirus type 2-B.
VIRUSES: ssDNA viruses; Circoviridae; Circovirus;
REFERENCE
Hamel,A.L. and Nayar,G.P.S.
Nucleotide sequence of four different isolates of porcine
circovirus detected in pigs with various clinical syndromes
JOURNAL Unpublished

```

```

REFERENCE 2 (bases 1 to 1768)
AUTHORS Hamel,A.L. and Nayar,G.P.S.
TITLE Direct Submission
JOURNAL Submitted (10-DEC-1998) Virology Laboratory, Manitoba Agriculture,
545 University Crescent, Winnipeg, Manitoba R3T 5S6, Canada
FEATURES
Location/Qualifiers
1..1768
/organism="porcine circovirus type 2-B"
/db_xref="taxon:85709"
/note="sequence obtained from several overlapping PCRs
using DNA extracted from lung, mesenteric lymph node and
tonsil of pig; similar to porcine circovirus sequence
presented in GenBank Accession Number AF027217; type-B
designation is based upon restriction endonuclease
digestion pattern"
repeat_region 13..35
/rpt_type=tandem
/rpt_unit=13..18
51..995
CDS
/codon_start=1
/product="putative Rep and coat protein"
/protein_id="AAD03086.1"
/db_xref="PID:g4106918"
/db_xref="GI:4106918"
/translation="MPSKNGSGPQPHKRWVTLNPSDEKRRKIRELPISLFDYFI
VGEENEGRTPHLOGFANFKVQKOTENKVKYLGARHIEKAKGTDOONKEYCSKEGN
LLIEGAPRSQGRSDLSSTAVSLLESGLVVAEOHPVTFVNRFLGAEELKVSCKM
QRDKNTNHHVIVGPPGCKSKAANFADPTTYWKPNNKWDGHEEVVDDFY
GWLWDLLRLCDRPLATYETGKTVFPLARSILITSNQTPLEWYSTAVPAVEALYR
RITSLVFWKNATEQSTEGGQFVTLSPCPPEPYEINY"
327..332
complement(357..671)
polyA_signal
CDS
/codon_start=1
/product="ORF-3"
/protein_id="AAD03087.1"
/db_xref="PID:g4106919"
/db_xref="GI:4106919"
/translation="WVTIPLVSRWFPVCGFRVCKISSPFAFTTPBPHNDVDIGLPI
TLHHPFAHQFQSPAEISDKRYRLLCGHQTALQGGTHSSRQVTPLSRSTF
NK"
complement(386..565)
CDS
/codon_start=1
/product="ORF-4"
/protein_id="AAD03088.1"
/db_xref="PID:g4106920"
/db_xref="GI:4106920"
/translation="MTWTLVFSRFCIFPLTFKSSASPRKFLTNVTGCCSATVRLPL
SNKVLTAVDRLRCP"
553..732
CDS
/codon_start=1
/product="ORF-12"
/protein_id="AAD03089.1"
/db_xref="PID:g4106921"
/db_xref="GI:4106921"
/translation="WSTSLWGLGVVVKANGLLILQTRKPHTGHNLETSGGVTWVKW
LLLMFMACRCNII"
complement(688..753)
CDS
/codon_start=1
/product="ORF-8"
/protein_id="AAD03094.1"
/db_xref="PID:g4106926"
/db_xref="GI:4106926"
/translation="MDIDHTVSVDPHTAASHKSHQ"
983..988
polyA_signal
CDS
complement(989..1033)
/codon_start=1
/product="ORF-11"
/protein_id="AAD03090.1"
/db_xref="PID:g4106922"
/db_xref="GI:4106922"
/translation="MNNKNHYEVIKKTQ"
1016..1177
CDS

```

```

/codon_start=1
/product="ORF-5"
/protein_id="AADO3091.1"
/db_xref="PID:g4106923"
/db_xref="GI:4106923"
/translation="MVEIILHGFKMGVFKIKFSELYIHGYTDIVVLVVFVTFERSAEA
YVHHSRGL"
polyA_signal
CDS
    complement(1022..1027)
    complement(1256..1735)
    note="similar to Porcine circovirus ORF-2 encoded by the
sequence presented in GenBank Accession Number AF07217."
/codon_start=1
/product="ORF-2"
/protein_id="AADO3085.1"
/db_xref="PID:g4106917"
/db_xref="GI:4106917"
/translation="MTYPRRYRRRHPRSHLGOILRRPWLPHRHYRWRRKNGI
FNRLSTFGYTKATVTTPSWAVDMRNFIDFVPPGGTNRKISVPPFFYIRKVK
VEFWPCSPITQDGRGVGSTAVILDDNEVTRATALTYPVYNSRHTIPQPFYSHR"
    complement(1522..1611)
    /codon_start=1
    /product="ORF-6"
    /protein_id="AADO3084.1"
    /db_xref="PID:g4106916"
    /db_xref="GI:4106916"
    /translation="MASSTPASPAPSDILSLPQSRPPGGRWT"
    1524..1631
    /codon_start=1
    /product="ORF-10"
    /protein_id="AADO3092.1"
    /db_xref="PID:g4106924"
    /db_xref="GI:4106924"
    /translation="MSTAQEGVTVALTVPKVRRRLKMPFFLLQR"
    1682..1741
    /codon_start=1
    /product="ORF-7"
    /protein_id="AADO3093.1"
    /db_xref="PID:g4106925"
    /db_xref="GI:4106925"
    /translation="MAAGAVSSSAVTPPWIRHS"
    join(1750..1768,1..13)
    note="putative replication site"
    join(1762..1768,1..2)
    note="similar to the nonanucleotide motif of Porcine
circovirus presented in GenBank Accession Numbers AF07217
and U49186"
BASE COUNT      448 a   360 c   493 g   467 t
ORIGIN

alignment_scores:
    Quality: 531.00      Length: 104
    Ratio: 5.257        Gaps: 0
    Percent Similarity: 97.115      Percent Identity: 94.231
alignment_block:
US-09-209-961-9 x AF112862/rev ..
Align seg 1/1 to reverse of: AF112862 from: 1 to: 1768
1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
671 ATGGTAACCATCCACCACTTTCTAGTAGGTGGTTCCAGTAGTGGTTT 622
17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
621 CCGGTCTGCAAAATAGCAGCCCATTCCTTTACCAACCCAGGTGC 572
34 roHisAsnGluValThrIleGlyPheProIleThrLeuLeuHisPhePro 50
571 CCCAATGACGTGCATGTGCTTCCATCAGCTCTCTGCTATTTCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67

```

```

|||||
521 GCTCACTTTCAAAAGTTCAGCCAGCCGCGGAAATTTCTGACAAACGTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlnGlyT 84
|||||
471 CAGGGTGTCTGCTGCAACGGTCACCAGACTCCCGCTCTCCAACAAGTA 422
84 hHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSerSer 100
|||||
421 CTCACAGCAGTAGACAGGTACCTCCGTTGTCCTTGAGATCAGAGAGCTCC 372
101 ThrPheAsnLys 104
|||||
371 ACATTCAATAAG 360
seq_name: gb_v1:AF154679
seq_documentation_block:
LOCUS      AF154679          1768 bp      DNA      circular      VRL      13-JUN-1999
DEFINITION Porcine Circovirus, complete genome.
ACCESSION  AF154679
NID        95052004
VERSION     AF154679.1  GI:5052004
KEYWORDS
SOURCE      porcine circovirus.
ORGANISM    Porcine circovirus.
REFERENCE   1 (bases 1 to 1768)
AUTHORS     Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE       Complete nucleotide sequences analysis of porcine circovirus
            outbreak in Taiwan
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 1768)
AUTHORS     Kuo,T.Y., Chiou,Y.C. and Lai,S.S.
TITLE       Direct Submission
JOURNAL     Submitted (27-MAY-1999) Veterinary Medicine, National Taiwan
            University, 142, Chousan Road, Taipei 106, Taiwan
FEATURES
            source
            1..1768
            /organism="porcine circovirus"
            /isolate="MLTW98"
            /db_xref="taxon:46221"
            /country="Taiwan"
            51..995
            /note="putative Rep protein"
            /codon_start=1
            /product="ORF1"
            /protein_id="AAD38398.1"
            /db_xref="PID:95052005"
            /db_xref="GI:5052005"
            /translation="MPSKKNGRSGPQPHKRVFTLNNPSEDERKKIRELPISLDFCFI
            VGEENEGRTPHLQGFANFYKOTFNKVKYFGARCHIEKAKGTDQONKEYCSKGN
            LLIECGAPRSQGRSDLSLSTAVSLTLESGSLVTAEOHPVTFVRNFRGLAELKVGKM
            OKRDKTNHVIVGPGCGKSKMAANFADPTTYKPKRWMDCYHGEEVVIDFY
            GWLPDDLRLCDDRYPLTVETKGTGTVPLARSILITSNQPLEWTSSTAVPAVEALYR
            RITSLVFWNKATEOSTEGGOFVLSPPCPEFFYEINY"
BASE COUNT      451 a   367 c   495 g   455 t
ORIGIN

alignment_scores:
    Quality: 522.00      Length: 104
    Ratio: 5.168        Gaps: 0
    Percent Similarity: 97.115      Percent Identity: 92.308
alignment_block:
US-09-209-961-9 x AF154679/rev ..
Align seg 1/1 to reverse of: AF154679 from: 1 to: 1768
1 MetValThrIleProProLeuValPheArgTrpPheProValCysGlyPh 17
671 ATGGTAACCATCCACCACTTTCTAGTAGGTGGTTTCCAGTAGTGGTTT 622

```

Ratio: 5.168 Gaps: 0
Percent Similarity: 97.115 Percent Identity: 92.308
alignment_block:
US-09-209-961-9 x AF166528/rev
Align seg 1/1 to reverse of: AF166528 from: 1 to: 1768
1 MetValThrIleProLeuValPheArgTrpPheProValCysGlyph 17
671 ATGTATACCAATCCACCACTTGTTCGAGGTGGTTTCCAGTATGTTGTT 622
17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
621 CCGGTCTGCAAAATTAGCAGCCCATTTTGTTCACACACCCAGGTGC 572
34 rOHISAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
571 CCCCAATGACGTGTACATTGTCTTCCAATCAGCTTCTGCATTTTCCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
521 GCTCATTTCAAAAGTTTCAGCCGCCCGGAAATTTCTGACAAACGTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
471 CAGGTGCTGCTCTGCAACGGTTCACGACTCCCGCTCTCCAACAAGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
421 CTCACAGCAGTAGACAGTCACTGCGTTTGTCTTGTGATCTAGAGCTCC 372
101 ThrPheAsnLys 104
||||:|||||
371 ACACTCGATAAG 360

17 eArgValCysLysIleSerSerProPheAlaPheThrThrProArgTrp 34
621 CCGGTCTGCAAAATTAGCAGCCCATTTGTTCACACACCCAGGTGC 572
34 rOHISAsnGluValTyrlleGlyPheProIleThrLeuLeuHisPhePro 50
571 CCCCAATGACGTGTACATTGTCTTCCAATCAGCTTCTGCATTTTCCC 522
51 AlaHisPheGlnLysPheSerGlnProAlaGluIlePheAspLysArgTy 67
521 GCTCATTTCAAAAGTTTCAGCCGCCCGGAAATTTCTGACAAACGTTA 472
67 rArgValLeuLeuCysAsnGlyHisGlnAsnProAlaLeuGlnGlyT 84
471 CAGGTGCTGCTCTGCAACGGTTCACGACTCCCGCTCTCCAACAAGTA 422
84 hrHisSerSerArgGlnValThrProLeuSerLeuArgSerArgSer 100
421 CTCACAGCAGTAGACAGTCACTGCGTTTGTCTTGTGATCTAGAGCTCC 372
101 ThrPheAsnLys 104
||||:|||||
371 ACACTCGATAAG 360
seq_name: gb_v1:AF166528
seq_documentation_block:
LOCUS AF166528 1768 bp mRNA VRL 19-AUG-1999
DEFINITION Porcine circovirus complete genome.
ACCESSION AF166528
NID g5739338
VERSION AF166528.1 GI:5739338
KEYWORDS
SOURCE porcine circovirus.
ORGANISM porcine circovirus
Virus; ssDNA viruses; Circoviridae; Circovirus.
REFERENCE 1 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Complete nucleotide sequences of porcine circovirus Tainan strand outbreak in Taiwan
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1768)
AUTHORS Yang, K.H., Lee, Y.F., Chao, D.S., Shieh, Y.C. and Lai, S.S.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1999) Veterinary Medicine, National Chia-Yi Institute of Technology, 300 Shei Fu Road, Lu Liao Li, Chia-Yi City 600, Taiwan
FEATURES
source
1..1768
/organism="porcine circovirus"
/virion
/strain="Tainan"
/db_xref="taxon:46221"
/country="Taiwan"
51..995
/codon_start=1
/product="putative Rep protein"
/protein_id="AAD50432.1"
/db_xref="PID:g5739339"
/db_xref="GI:5739339"
/translation="MPSKNGSGPQPKRWVFTLNNPSEDERKKIRELPISLFDYFI
VGEENEGRTRPHLOGFANFKQFNKVKYFGARCHEKAKGTDQONKEYCSKEGN
LLIEGAPRSQQRSDLSVSTLLESGLVTAEOHPVTFVNFRLAEALLKVSQKM
QKRDKNVHVTVPGGSKSAANFADPTTYKPPRNKWDGHHGEVYVDDFY
GHLPWDDLRLCLDRPLVETRGKTVPLFARLSILITSNOTPLEWYSSVAPVAEALYR
RTISLVEWKNATEQSTEESGQFVTLSPCPPEFYINY"
BASE COUNT 453 a 367 c 492 g 456 t
ORIGIN
alignment_scores:
Quality: 522.00 Length: 104

WILEY
***** (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed Dec 22 10:25:26 1999; MasPar time 12.23 Seconds
Tabular output not generated. 102.575 Million cell updates/sec

Title: >US-09-209-961-20
Description: (1-59) from US09209961.pep
Perfect score: 471
Sequence: 1 MYTSLWGLGVKRWANGLLIL.....MVKKLLMTFMAGCRGMIV 59

Scoring table: PAM 150
Gap 11

Searched: 170751 seqs, 2126608 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq35
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23
24:part24 25:part25 26:part26 27:part27 28:part28
29:part29 30:part30 31:part31 32:part32 33:part33
34:part34 35:part35 36:part36 37:part37 38:part38
39:part39

Statistics: Mean 26.466; Variance 97.683; scale 0.271

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	78	16.6	365	13	Prostaglandin-EP3-9 r	2.36e+01
2	78	16.6	388	13	Prostaglandin-EP3-21	2.36e+01
3	78	16.6	390	13	Prostaglandin-EP3-1p	2.36e+01
4	78	16.6	393	32	Human prostaglandin E	2.36e+01
5	78	16.6	402	32	Human EP3-V receptor	2.36e+01
6	78	16.6	427	6	benzene dioxygenase p	2.36e+01
7	78	16.6	441	32	Amino acid sequence o	2.36e+01
8	78	16.6	672	3	Sequence of thermosta	2.36e+01
9	78	16.6	672	2	Thermostable beta-gal	2.36e+01
10	77	16.3	534	2	Sequence of Ricin com	2.87e+01
11	77	16.3	4630	23	Streptomyces venezuel	2.87e+01
12	76	16.1	540	24	Castor oil plant aggl	3.49e+01
13	76	16.1	540	23	R. communis agglutini	3.49e+01
14	75	15.9	189	34	TNF related endotheli	4.24e+01
15	75	15.9	305	29	H. pylori ORF 02ae116	4.24e+01
16	75	15.9	332	29	H. pylori ORF 02ae116	4.24e+01

17 74 15.7 231 3 260030 Sequence of ricin B c
18 74 15.7 251 1 R11329 The ricin B chain fro
19 74 15.7 251 1 P11867 Amino acid sequence o
20 74 15.7 251 1 P90837 Sequence encoded by i
21 74 15.7 262 1 P81852 Ricin B mutein with d
22 74 15.7 262 1 P81849 Ricin B mutein with d
23 74 15.7 262 1 P90020 B chain of ricin D
24 74 15.7 262 1 P82853 Ricin B mutein with d
25 74 15.7 262 1 P82896 Ricin B mutein with d
26 74 15.7 262 1 P82865 Generic sequence for
27 74 15.7 262 1 P82851 Ricin B mutein with d
28 74 15.7 262 1 P82850 Ricin B mutein with d
29 74 15.7 306 35 W11512 Helicobacter polypt
30 74 15.7 317 3 P60507 Human apolipoprotein
31 74 15.7 317 16 R86791 Human apolipoprotein
32 74 15.7 317 3 P51204 Human apolipoprotein
33 74 15.7 562 1 P90079 Ricin D
34 74 15.7 565 3 P50166 Sequence of preproric
35 74 15.7 565 3 P60240 Preproricin.
36 74 15.7 574 2 P70325 Sequence of Ricinus c
37 74 15.7 574 1 P94793 DNA sequence of ricin
38 74 15.7 576 2 P70326 Sequence of Ricinus c
39 74 15.7 576 26 W25787 Castorbean ricin.
40 73 15.5 73 3 R12605 SIB 136 intestinal mu
41 73 15.5 102 14 W74204 Human death associate
42 73 15.5 102 36 W71366 Death associated prot
43 72 15.3 236 26 P60039 Sequence of ricin B c
44 72 15.3 236 2 P91888 Sequence of ricin B c
45 72 15.3 584 2 R06520 Microspore-specific c

ALIGNMENTS

RESULT 1
ID R69518 standard; Protein; 365 AA.
AC R69518;
DE 02-SEP-1994 (first entry)
KW Prostaglandin-EP3-9 receptor;
AW human uterus prostaglandin-EP3-9 receptor;
KW prostaglandin-EP3 receptor-agonist;
KW prostaglandin-EP3 receptor-antagonist.
CS Homo sapiens.
FH Key
FT modified_site 18 Location:Qualifiers
FT modified_site 36 /note= "N-glycosylation site"
FT modified_site 217 /note= "N-glycosylation site"
FT modified_site 308 /note= "N-glycosylation site"
FT region 338..345 /note= "N-glycosylation site"
FT W095JJ552-A.
FT 05-JAN-1995
FT 09-JUN-1994; CAC020.
FT 25-JUN-1993; US-083746.
FT (MERL) MERCK FROSST CANADA INC.
FT WPI: 95-052012/07.
FT N-P-DB; Q81951.
FT New prostaglandin EP3 receptors and DNA - used partic. to
FT identify modulators of prostaglandin receptor activity for
FT treatment of diseases.
FT Clin 2; Page 43-5; 64pp; English.
FT The sequence represents a human uterus prostaglandin-EP3-9
FT receptor (mol.wt. 40,507). The sequence contains 4 conserved
FT N-glycosylation sites in putative extracellular regions, and a
FT conserved region (R69515) in transmembrane region-VII common to
FT other EP3 receptors. Conserved cys residues are found in
FT exofacial loops 1 and 2, and a conserved Arg found in all
FT eicosanoid receptors is found in transmembrane region-VII. The
FT receptor may be used to identify receptor-agonists and

Figure 1. The effect of the concentration of the *Agrobacterium* suspension on the transformation efficiency of *Agrobacterium* strains.

CC preventive and treating agent for inflammation.
SQ Sequence 393 AA;

Query Match 16.6%; Score 78; DB 32; Length 393;
Best Local Similarity 31.3%; Pred. No. 2.36e+01;
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 lairaphyashmktravllgwlavlaf 188
QY 20 LQTRKPH-TGNELKTSGGMTVMVKWLLMTF 50

RESULT 5

ID W57410 standard; Protein: 402 AA.
AC W57410;
DT 19-AUG-1998 (first entry)
DE Human EP3-V receptor.
KW Prostaglandin E2 receptor; EP3-V receptor; human; treatment;
OS Homo sapiens.
PN J10113185-A.
PF 06-MAY-1998.
PD 14-OCT-1996; 291150.
PR 14-OCT-1996; JP-291150.
PA (ONOX) ONO PHARM CO LP.
DR WPI: 98-315474/28.
DN N-PSDB: V29610.
PT New human prostaglandin EP3 receptor(s) - useful for treatment and prevention of, e.g. inflammation.
PS Claim 2; Pages 16-17; 27pp; Japanese.
CC This represents a human EP3-V receptor. A replication or expression vector comprising cDNA sequences encoding EP-3V or EP3-3VI can be used to transform a host cell. The host cell is cultured and the polypeptides can be recovered from the culture medium. The polypeptides combine specifically with a prostaglandin PGE2 receptor and can be used as a preventive and treating agent for inflammation.
SQ Sequence 402 AA;

Query Match 16.6%; Score 78; DB 32; Length 402;
Best Local Similarity 31.3%; Pred. No. 2.36e+01;
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 lairaphyashmktravllgwlavlaf 188
QY 20 LQTRKPH-TGNELKTSGGMTVMVKWLLMTF 50

RESULT 6

ID R32085 standard; Protein: 427 AA.
AC R32085;
DT 07-JUN-1993 (first entry)
DE Benzene dioxygenase polypeptide I.
KW Benzene; dioxygenase; cysbenzene glycol; transformation; BDO.
OS Pseudomonas aeruginosa.
PN J05003785-A.
PD 14-JAN-1993.
PF 30-OCT-1991; 284795.
PR 16-APR-1991; JP-084170.
PA (SAOC) MERCIAN CORP.
DR WPI: 93-054795/07.
DN N-PSDB: Q36395.
PT Microbe DNA - contg. benzene di-oxygenase gene, cis:benzene glycol dehydrogenase gene and catechol 2,3-oxygenase gene
PS Claim 12; Page 4-5; 20pp; Japanese.
CC The sequences given in R32085-89 represent five regions of a benzene dioxygenase (BDO) polypeptide. Cysbenzene glycol can be produced CC effectively from benzene using transformants prepared using the DNA CC encoding these fragments.
SQ Sequence 427 AA;

Query Match 16.6%; Score 78; DB 6; Length 427;
Best Local Similarity 38.6%; Pred. No. 2.36e+01;
Matches 17; Conservative 8; Mismatches 16; Indels 3; Gaps 3;

Db 60 lgheshypetgdfiatymgedpvmvirkdksikvflnqcrqm 103
QY 17 LLILQTRKPH-TGNELKTSGGMTVMVKWLLMTF 57

RESULT 7

ID W28236 standard; Protein: 441 AA.
AC W28236;
DT 07-SEP-1998 (first entry)
DE Amino acid sequence of a mercuric reductase.
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
KW toxic shock syndrome; mercuric reductase.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT Misc difference: 16 /note= "Residue not specified"
PN WO9730070-A1
PD 21-AUG-1997
PF 19-FEB-1997; U02318.
PR 20-FEB-1996; US-011888.
PS (SMIK) SMITHKLINE BEECHAM CORP.
PI Black Mt, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO,
PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;
DR WPI: 97-124969/39.
DN N-PSDB: T84151.
PT Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used to isolate antimicrobial compounds, and in vaccines against S. aureus infection
PS Claim 1; Pages 547-548; 939pp; English.

The present sequence represents a Staphylococcus aureus protein, that, based on homology with another Staphylococcus aureus protein, is believed to be a mercuric reductase. The DNA sequence was isolated from a library of clones of S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used in the construction of ribozymes and antisense sequences to control the expression of Staphylococcal genes. The DNA sequence is also useful as a source of regulatory elements for the control of bacterial gene expression. The present protein may be used to produce vaccines to enable a host to produce specific antibodies with antibacterial action. These vaccines and antibodies would protect a host against invasion by S. aureus, and conditions relating to Staphylococcal infection, e.g. Staphylococcal food poisoning, scaled skin syndrome, and toxic shock syndrome.
SQ Sequence 441 AA;

Query Match 16.6%; Score 78; DB 32; Length 441;
Best Local Similarity 26.7%; Pred. No. 2.36e+01;
Matches 16; Conservative 14; Mismatches 28; Indels 2; Gaps 2;

Db 294 lyievngqevleadqvlvratgrkptet-lnlesagvtgkgevltnelylqtsmriy 352
QY 1 MYTSLMGLGVYKANGLLIQLTRKPH-TGNELKTSGGMTVMVKWLLMTF-FMAGCRGMII 59

RESULT 8

ID P60219 standard; Protein: 672 AA.
AC P60219;
DT 24-AUG-1991 (first entry)
DE Sequence of thermostable beta-galactosidase.
KW Low lactose milk; galactose; lactose intolerance; food processing.
OS Bacillus sphaerothermophilus.
PN EP-176971-A.
PD 09-APR-1986.
PF 29-SEP-1984; JP-202965.
PR 24-MAY-1986; JP-110565.
PS (WAKA) WAKA-MOTO PHARM KK.
PI Hirata H, Ikada H, Negoro S;
DR WPI: 86-G5363/15.
DN N-PSDB: N60150.
PT Thermostable beta-galactosidase from bacillus subtilis

DT 02-DEC-1997 (first entry)
 DE Castor oil plant agglutinin inactive precursor.
 KW Maize; proRIP; ribosome inactivating protein; alpha; beta subunit;
 KW internal linker; barley translation inhibitor; Trichosanthin;
 KW Ricin A-chain; Abrin-A-chain; Saporin; SLT-1; Luffin A; MAP;
 KW Ricinus communis agglutinin; Momordin; PAP-S; Luffin-B; Bianthrin 30;
 KW therapeutic toxin; tumour cell targeted; protein synthesis inhibitor;
 KW post-translational modification; cancer; neoplasia; HIV; AIDS;
 KW human immunodeficiency virus; acquired immune deficiency syndrome.
 OS Ricinus communis.
 PN US5646026-A.
 PD 08-JUN-1997. 535636.
 PE 11-JUN-1990. 535636.
 PR 09-DEC-1992; US-987927.
 PR 11-JUN-1990; US-535636.
 PR 26-JAN-1995; US-378761.
 PR 07-JUN-1995; US-485286.
 PA (DOWC) DOWELANCO.
 PI Hey TD, Morgan AER, Walsh TA.
 DR WPI; 97-362934/33.
 PT DNA encoding pro-ribosome inactivating proteins - inactive
 PT precursors of ribosome inactivating proteins; can be expressed in
 PT eukaryotic cells without causing cell death.
 PS Claim 4; Column 121-124; 186pp; English.
 CC W25143 shows a castor oil plant agglutinin protein which was engineered
 CC to contain a selectively removable internal peptide linker sequence
 CC separating the alpha and beta units. When separated the two units regain
 CC activity and are capable of inactivating eukaryotic ribosomes and hence
 CC preventing protein production. Many different ribosome inhibitory
 CC proteins (RIPs) may be produced with an internal linker including
 CC maize RIP, Trichosanthin, Ricin A-chain, Abrin-A-chain and
 CC Saporin. The RIPs can be used in the construction of therapeutic
 CC toxins targeted to specific cells such as tumour cells via the
 CC attachment of a targeting polypeptide, e.g. a monoclonal antibody.
 CC A further use is in HIV therapy (see US4869903). There is interest
 CC in expressing RIP recombinantly in host eukaryotic cells, because of
 CC the capacity to provide correct post-translational processing. However,
 CC RIPs effectively inhibit protein synthesis in eukaryotic cells resulting
 CC in cell death. Since the inactive RIP proteins are not cytotoxic to
 CC eukaryotic cells, they can be recombinantly expressed in such cells and
 CC then converted to active RIP proteins.
 SQ Sequence 540 AA;

Query Match 16.1%; Score 76; DB 24; Length 540;
 Best Local Similarity 17.1%; Pred. No. 3.49e+01;
 Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 indnrtlnprsglvlaatsngskltvtqtniyavsgqw 410
 QY 5 LMGHLGVKVRANGLLILQTRKPHGTGNHLKTSGGWTVMKWL 45
 || : : : : : || : : : : : || : : : : : || : : : : : ||

RESULT 13
 ID W21706 standard; Protein; 540 AA.
 AC W21706;
 DT 26-SEP-1997 (first entry)
 DE R. communis agglutinin.
 KW pro-Ribosome inactivating protein; proRIP; peptide linker; cancer;
 KW inactivation; eukaryotic ribosome; alpha fragment; beta fragment;
 KW inhibitor; protein synthesis; N-glycosidase; glycosidic bond; liver;
 KW rat; ribosomal 28S RNA; cellular proliferation; HIV-infected T cell.
 OS Ricinus communis.
 FH Key Location/Qualifiers
 FT region 152..162
 FT /note= "Position of possible insertion of internal
 FT peptide linker sequence"
 PN US5635384-A.
 PD 03-JUN-1997. 535636.
 PE 11-JUN-1990. 535636.
 PR 11-JUN-1990; US-535636.
 PR 09-DEC-1992; US-987927.
 PR 26-JAN-1995; US-378761.
 PA (DOWC) DOWELANCO.

PI Hey TD, Morgan AER, Walsh TA.
 DR WPI; 97-309331/28.
 PT Inactive precursor of maize ribosome-inactivating protein - also
 PT chimeric ribosome-inactivating protein precursors containing
 PT internal linker sequences
 PS Claim 2; Column 119-124; 121pp; English.
 CC The sequences given in W21698-710 represent Ribosome Inactivating
 CC proteins (RIPs); which may be used in the construction of the
 CC proRIP of the invention. The proRIP has a selectively removable,
 CC internal peptide linker. The precursor sequence is incapable of
 CC inactivating eukaryotic ribosomes, but can be converted by removal
 CC of the linker into a protein having alpha and beta fragments and being
 CC capable of inactivating eukaryotic ribosomes. RIPs are potent
 CC inhibitors of eukaryotic protein synthesis. They possess a highly
 CC specific N-glycosidase activity which cleaves the glycosidic bond of
 CC adenine 4324 of rat liver ribosomal 28S RNA. RIP's selectively inhibit
 CC cellular proliferation of cells, e.g. cancer cells and HIV-infected T
 CC cells. The inactive proRIP proteins make it possible to provide protein
 CC synthesis inhibitors with uses in practical and improved ways not before
 CC possible. The RIP can be used to make cytotoxic conjugates.
 SQ Sequence 540 AA;

Query Match 16.1%; Score 76; DB 23; Length 540;
 Best Local Similarity 17.1%; Pred. No. 3.49e+01;
 Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 indnrtlnprsglvlaatsngskltvtqtniyavsgqw 410
 QY 5 LMGHLGVKVRANGLLILQTRKPHGTGNHLKTSGGWTVMKWL 45
 || : : : : : || : : : : : || : : : : : || : : : : : ||

RESULT 14
 ID W29746 standard; Protein; 189 AA.
 AC W29746;
 DT 27-OCT-1998 (first entry)
 DE TNF related endothelium proliferative agent protein 2.
 KW TNF; endothelium proliferative agent; TREPA; wound healing; cancer;
 KW tissue grafting; vascularisation; apoptosis; autoimmune; birth control.
 OS Homo sapiens.
 PN WQ9835061-A2.
 PD 13-AUG-1998.
 PR 12-FEB-1998; U02859.
 PR 10-FEB-1998; US-021706.
 PR 12-FEB-1997; US-798692.
 PA (ABBO) ABBOTT LAB.
 PI Wiley SR;
 DR WPI; 98-447255/38.
 PT Detecting nucleic acid encoding TREPA - useful for diagnosis and
 PT treatment of autoimmune disease, tumours and inflammation
 PS Claim 16; Page 125-6; 142pp; English.
 CC The TNF-related endothelium proliferative agent (TREPA), or its
 CC activators or agonists, are used to treat a deficit of TREPA, e.g. to
 CC promote wound healing or tissue grafting, by promoting vascularisation,
 CC also to induce apoptosis for treating cancer and eliminating autoreactive
 CC T cells, as an adjunct to cancer chemotherapy or antiviral treatment.
 CC TREPA peptides can also be used to target cytotoxic agents or for
 CC affinity isolation of the corresponding receptor, the nucleic acid for
 CC which can be used to transform tumour cells to render them more
 CC responsive to TREPA and to screen for TREPA mimics.
 CC Ribozymes; antisense RNA, antibodies or peptides, are used to treat
 CC TREPA-associated diseases, e.g. tumours and metastases (by inhibiting
 CC vascularisation), inflammation or a wide range of autoimmune conditions,
 CC conditions involving abnormal stimulation of epithelial cells (e.g.
 CC atherosclerosis), for birth control (inhibiting ovulation and placental
 CC formation) or other angiogenic conditions (e.g. ulcers).
 SQ Sequence 189 AA;

Query Match 15.9%; Score 75; DB 34; Length 189;
 Best Local Similarity 45.0%; Pred. No. 4.24e+01;
 Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

Db 1 mymmtitwllpmvllaayc 20
 || : : : : : || : : : : : || : : : : : || : : : : : ||

QY 37 MVTMVKKW-LLMTFNAG-C 54

RESULT 15

ID W55451 standard; Protein; 305 AA.
 AC W55451;
 DT 24-JUN-1998 (first entry)
 DE H. pylori ORF 02aell612_23598175_fl_2 secreted protein.
 KW Cytoplasmic; vaccine; prevention; treatment; infection; envelope;
 KW identification; binding compound; bacteria; life cycle; activator;
 KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis.
 OS Helicobacter pylori.
 PN W09737044-AL.
 PD 09-OCT-1997.
 PF 27-MAR-1997; U05223.
 PR 06-DEC-1996; US-761318.
 PR 29-MAR-1996; US-625811.
 PR 02-APR-1996; US-758731.
 PR 25-OCT-1996; US-736905.
 PR 28-OCT-1996; US-738859.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Smith D;
 DR WPI; 97-503122/46.
 DR N-PSDB; V24860.
 PT Helicobacter pylori nucleic acid sequences and encoded
 PT polypeptide(s) - useful in vaccines to treat or prevent H. pylori
 PT infection and for diagnosis of H. pylori infection
 PS Claims 14, 94; Page 658; 1145pp; English.
 CC This sequence is a H. pylori secreted protein.
 CC The protein may be used in a vaccine to prevent or treat H. pylori
 CC infection or to identify H. pylori polypeptide binding compounds.
 CC useful as potential H. pylori life cycle activators or inhibitors. The
 CC DNA and probes derived from it may be used for the identification of
 CC H. pylori in a sample and the diagnosis of H. pylori infection. Nucleic
 CC acid sequences complementary to the DNA act as antisense sequences and
 CC can be used to prevent the translation of H. pylori mRNA. Antibodies
 CC against the protein can be used in immunoassays to evaluate the abundance
 CC and distribution of H. pylori-specific antigens. The genomic sequence of
 CC H. pylori (ATCC 53679) was determined from overlapping contigs generated
 CC by mechanically shearing the bacterial DNA. The sequences were analysed
 CC for ORF of at least 180 nucleotides, and the predicted coding regions
 CC defined by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from various ORF
 CC were analysed for significant homology to other known or exported
 CC membrane proteins. Having identified and determined the sequences of
 CC interest, particular regions can be isolated from H. pylori by PCR
 CC amplification for recombinant polypeptide production, e.g. in E. coli
 CC hosts.
 SQ Sequence 305 AA;

Query Match 15.9%; Score 75; DB 29; Length 305;
 Best Local Similarity 42.9%; Pred. No. 4.24e+01;
 Matches 9; Conservative 6; Mismatches 5; Indels 1; Gaps 1;

Db 1 mlkswtkkfwllflmscfcg 21

|: |||:: |||: |
 QY 37 MVTM-VKKLLMTFNAGCRG 56

Search completed: Wed Dec 22 10:25:43 1999
 Job time : 17 secs.

W O R L D

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (C) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:27:23 1999; MasPar time 3.60 Seconds
Tabular output not generated. 195.463 Million cell updates/set

Title: >US-09-209-961-20
Description: (1-59) from US09209961.pep
Perfect Score: 471
Sequence: 1 MYTSLGHLGVKXANGLLIL.....MVKKLLMTFMAGCRGMIV 59

Scoring table: PAM 150
Gap 11

Searched: 122461 seqs, 11912985 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:PC9_COMB 4:backfiles1

Statistics: Mean 24.934; Variance 98.568; scale 0.253

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	76	16.1	540	1	US-08-485- Sequence 77, Applicati	2.35e+01
2	76	16.1	540	1	US-08-378- Sequence 77, Applicati	2.35e+01
3	74	15.7	220	2	US-08-726- Sequence 29, Applicati	3.41e+01
4	74	15.7	317	1	US-07-709- Sequence 2, Applicati	3.41e+01
5	72	15.3	414	1	US-08-002- Sequence 19, Applicati	4.93e+01
6	72	15.3	414	1	US-08-002- Sequence 17, Applicati	4.93e+01
7	72	15.3	414	1	US-08-002- Sequence 13, Applicati	4.93e+01
8	72	15.3	484	1	US-08-030- Sequence 8, Applicati	4.93e+01
9	71	15.1	529	2	US-08-496- Sequence 2, Applicati	5.93e+01
10	70	14.9	510	1	US-08-278- Sequence 4, Applicati	7.12e+01
11	69	14.6	361	2	US-08-685- Sequence 4, Applicati	8.54e+01
12	69	14.6	361	1	US-08-390- Sequence 2, Applicati	8.54e+01
13	69	14.6	365	2	US-08-390- Sequence 2, Applicati	8.54e+01
14	69	14.6	365	2	US-08-685- Sequence 2, Applicati	8.54e+01
15	69	14.6	532	3	PCR-US93-0 Sequence 6, Applicati	8.54e+01
16	69	14.6	532	1	US-08-481- Sequence 6, Applicati	8.54e+01
17	68	14.4	235	1	US-07-932- Sequence 1, Applicati	1.02e+02
18	68	14.4	528	2	US-08-466- Sequence 2, Applicati	1.02e+02
19	68	14.4	528	2	US-08-700- Sequence 2, Applicati	1.02e+02
20	67	14.2	969	2	US-08-284- Sequence 2, Applicati	1.22e+02
21	67	14.2	969	3	PCR-US93-0 Sequence 2, Applicati	1.22e+02
22	67	14.2	3200	2	US-08-477- Sequence 8, Applicati	1.22e+02
23	66	14.0	181	2	US-08-353- Sequence 74, Applicati	1.46e+02

24	66	14.0	212	2	US-08-563- Sequence 2, Applicatio	1.46e+02
25	66	14.0	216	2	US-08-713- Sequence 7, Applicatio	1.46e+02
26	66	14.0	292	2	US-08-353- Sequence 11, Applicati	1.46e+02
27	66	14.0	349	1	US-08-459- Sequence 33, Applicati	1.46e+02
28	66	14.0	349	1	US-08-197- Sequence 33, Applicati	1.46e+02
29	66	14.0	349	1	US-08-459- Sequence 33, Applicati	1.46e+02
30	66	14.0	353	1	US-08-197- Sequence 43, Applicati	1.46e+02
31	66	14.0	353	1	US-08-459- Sequence 43, Applicati	1.46e+02
32	66	14.0	353	1	US-08-459- Sequence 43, Applicati	1.46e+02
33	66	14.0	894	1	US-08-486- Sequence 4, Applicatio	1.46e+02
34	66	14.0	894	1	US-08-117- Sequence 4, Applicatio	1.46e+02
35	65	13.8	251	2	US-08-602- Sequence 36, Applicati	1.75e+02
36	65	13.8	732	3	PCR-US93-1 Sequence 2, Applicatio	1.75e+02
37	65	13.8	884	2	US-08-982- Sequence 2, Applicatio	1.75e+02
38	65	13.8	884	2	US-08-465- Sequence 2, Applicatio	1.75e+02
39	64	13.6	42	2	US-08-472- Sequence 9, Applicatio	2.09e+02
40	64	13.6	291	1	US-08-467- Sequence 11, Applicati	2.09e+02
41	64	13.6	291	3	PCR-US96-0 Sequence 11, Applicati	2.09e+02
42	64	13.6	1408	2	US-08-612- Sequence 2, Applicatio	2.09e+02
43	64	13.6	1565	4	5352450-2 Patent No. 5352450.	2.09e+02
44	64	13.6	2183	1	US-08-348- Sequence 7, Applicatio	2.09e+02
45	64	13.6	3077	4	5223423-2 Patent No. 5223423.	2.09e+02

ALIGNMENTS

RESULT 1
ID US-08-485-286-77 STANDARD; PRT; 540 AA.
XX AC xxxxxx
YA
XX
XX
DE
XX

Sequence 77, Application US/08485286

Sequence 77, Application US/08485286

Patent No. 5646026

Patent No. 5646026 5646119

GENERAL INFORMATION:

APPLICANT: WALSH, TERENCE A

APPLICANT: HEY, TIMOTHY D

APPLICANT: MORGAN, ALICE ER

TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE

TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD

TITLE OF INVENTION: USING

NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:

ADDRESSEE: ANDREA T. BORUCKI

STREET: 19330 ZIONSVILLE ROAD

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,286

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/378761

FILING DATE: 26-JAN-1995

ATTORNEY/AGENT INFORMATION:

NAME: BORUCKI, ANDREA T

REGISTRATION NUMBER: 33651

REFERENCE/DOCKET NUMBER: 38272B

TELECOMMUNICATION INFORMATION:

TELEPHONE: (317) 337-4846

INFORMATION FOR SEQ ID NO: 77:

SEQUENCE CHARACTERISTICS:

CC LENGTH: 540 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 540 AA; 60168 MW; 1511677 CN;

Query Match 16.1%; Score 76; DB 1; Length 540;
Best Local Similarity 17.1%; Pred. No. 2.35e+01;
Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 IWDNRTIINPRSGVLATSGNSGKLVQVNIYAVSOGWL 410
Qy 5 LWGHLGVVKGALLILQTRKPHGTGNHLATSGGWTWVKWL 45

RESULT 2

ID US-08-378-761A-77 STANDARD; PRT; 540 AA.
XX xxxxxx

Sequence 77, Application US/08378761A

Sequence 77, Application US/08378761A

Patent No. 5635384

GENERAL INFORMATION:

APPLICANT: WALSH, TERENCE A

APPLICANT: HEY, TIMOTHY D

APPLICANT: MORGAN, ALICE ER

TITLE OF INVENTION: RIBOSOME-INACTIVATING PROTEINS, INACTIVE

TITLE OF INVENTION: PRECURSOR FORMS THEREOF, A PROCESS FOR MAKING A METHOD OF

TITLE OF INVENTION: USING

NUMBER OF SEQUENCES: 81

CORRESPONDENCE ADDRESS:

ADDRESSEE: ANDREA T. BORUCKI

STREET: 9330 ZIONSVILLE ROAD

CITY: INDIANAPOLIS

STATE: IN

COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/378,761A

FILING DATE: 26-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BORUCKI, ANDREA T

REGISTRATION NUMBER: 33651

REFERENCE/DOCKET NUMBER: 38272B

TELEPHONE: (317) 337-4846

INFORMATION FOR SEQ ID NO: 77:

SEQUENCE CHARACTERISTICS:

LENGTH: 540 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE 540 AA; 60168 MW; 1511677 CN;

Query Match 16.1%; Score 76; DB 1; Length 540;
Best Local Similarity 17.1%; Pred. No. 2.35e+01;
Matches 7; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

Db 370 IWDNRTIINPRSGVLATSGNSGKLVQVNIYAVSOGWL 410
Qy 5 LWGHLGVVKGALLILQTRKPHGTGNHLATSGGWTWVKWL 45

QTRKPH⁺TGNHLK⁺TS⁺GMV⁺TMVKKWLL-LMTFMAGC 54

CC FILING DATE: 23-FEB-1993
CC APPLICATION NUMBER: JP 036580-1992
CC FILING DATE: 24-FEB-1992
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: JP 064889-1992
CC FILING DATE: 23-MAR-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Fordis, Jean B.
CC REGISTRATION NUMBER: 32,984
CC REFERENCE/DOCKET NUMBER: 04221-0020-000000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 202-408-4000
CC TELEFAX: 202-408-4400
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 365 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
CC SEQUENCE 365 AA; 40077 MW; 704770 CN;
CC
CC Query Match 14.6%; Score 69; DB 1; Length 365;
CC Best Local Similarity 34.4%; Pred. No. 8.54e+01;
CC Matches 11; Conservative 9; Mismatches 10; Indels 2; Gaps 2;
CC
Db 134 LAIRAPHYASHMKTRATPV-LLGVWLSVLAF 164
QY 20 LQTRKPH-TGNHLKTSGGMTVMVKWLLMTF 50
RESULT 14
ID US-08-685-945B-2 STANDARD; PRT; 365 AA.
XX AC xxxxxx
XX DT
XX DE
XX SE
XX Sequence 2, Application US/08685945B
XX Patent No. 5804415
XX GENERAL INFORMATION:
XX APPLICANT: Ichikawa, Atsushi
XX APPLICANT: Narumiya, Shuh
XX TITLE OF INVENTION: Prostaglandin E Receptors, Their DNA and
XX NUMBER OF SEQUENCES: 6
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
XX STREET: 1300 I Street, N.W., Suite 700
XX CITY: Washington
XX STATE: D.C.
XX COUNTRY: USA
XX ZIP: 20005-3315
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: Patentin Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: US/08/685,945B
XX FILING DATE: 22-JUL-1996
XX CLASSIFICATION: 435
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: US 08/024179
XX FILING DATE: 23-FEB-1993
XX CLASSIFICATION: 435
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: JP 036580-1992
XX FILING DATE: 24-FEB-1992
XX PRIOR APPLICATION DATA:
XX APPLICATION NUMBER: JP 064889-1992
XX FILING DATE: 23-MAR-1992
XX ATTORNEY/AGENT INFORMATION:

NAME: Fordis, Jean B.
REGISTRATION NUMBER: 32,984
REFERENCE/DOCKET NUMBER: 04221-0020-02000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 365 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE 365 AA; 40077 MW; 704770 CN;
Query Match 14.6%; Score 69; DB 2; Length 365;
Best Local Similarity 34.4%; Pred. No. 8.54e+01;
Matches 11; Conservative 9; Mismatches 10; Indels 2; Gaps 2;
Db 134 LAIRAPHYASHMKTRATPV-LLGVWLSVLAF 164
QY 20 LQTRKPH-TGNHLKTSGGMTVMVKWLLMTF 50
RESULT 15
ID PCT-US95-05467-6 STANDARD; PRT; 532 AA.
XX AC xxxxxx
XX DT
XX DE
XX SE
XX Sequence 6, Application PC/TUS9505467
XX GENERAL INFORMATION:
XX APPLICANT:
XX APPLICANT:
XX TITLE OF INVENTION: MORPHOGENIC PROTEIN-SPECIFIC CELL
XX TITLE OF INVENTION: SURFACE RECEPTORS AND USES THEREFOR
XX NUMBER OF SEQUENCES: 15
XX CORRESPONDENCE ADDRESS:
XX ADDRESSEE: PATENT ADMINISTRATOR, TESTA, HURWITZ &
XX ADDRESS: THIBEAULT
XX STREET: 53 STATE STREET
XX CITY: BOSTON
XX STATE: MA
XX COUNTRY: USA
XX ZIP: 02109
XX COMPUTER READABLE FORM:
XX MEDIUM TYPE: Floppy disk
XX COMPUTER: IBM PC compatible
XX OPERATING SYSTEM: PC-DOS/MS-DOS
XX SOFTWARE: Patentin Release #1.0, Version #1.25
XX CURRENT APPLICATION DATA:
XX APPLICATION NUMBER: PCT/US95/05467
XX FILING DATE:
XX CLASSIFICATION:
XX ATTORNEY/AGENT INFORMATION:
XX NAME: FITCHER, EDMUND R.
XX REGISTRATION NUMBER: 27,829
XX REFERENCE/DOCKET NUMBER: CRP-097PC
XX TELECOMMUNICATION INFORMATION:
XX TELEPHONE: (617) 248-7000
XX TELEFAX: (617) 248-7100
XX INFORMATION FOR SEQ ID NO: 6:
XX SEQUENCE CHARACTERISTICS:
XX LENGTH: 532 amino acids
XX TYPE: amino acid
XX TOPOLOGY: linear
XX MOLECULE TYPE: protein
XX SEQUENCE 532 AA; 60201 MW; 1453753 CN;
Query Match 14.6%; Score 69; DB 3; Length 532;
Best Local Similarity 47.1%; Pred. No. 8.54e+01;

Tue Jan 4 11:52:15 2000

US-09-209-961-20.rai

Page 8

Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

Db 152 RWLVLLISMVCIITAMI 168

:|:|: || |:|:

Oy 43 KWLLIMTFMAGCR-GMI 58

Search completed: Wed Dec 22 10:27:30 1999
Job time : 7 secs.

```

mpsrch_pp      protein - protein database search, using Smith-Waterman algorithm
Run on:        Wed Dec 22 10:26:00 1999;  Maspar time 6.68 Seconds
               ---353.834 Million cell updates/sec
Tabular output not generated.

```

```

Scoring table:      PAM 150
                   Gap 11
Searched:          122810 seqs, 40068593 residues
Post-processing:    Minimum Match 0%
                   Listing first 45 summaries
Database:           pir60
                   1:pir1 2:pir2 3:pir3 4:pir4
Statistics:         Mean 36.673; Variance 66.415; scale 0.552

```

SUMMARIES

Result No.	Query		DB	ID	Description	Pred. No.
	Score	Match Length				
1	95	20.2	593	2	S72769	BL496_Fl_14 protein - 1.57e-02
2	94	20.0	591	2	F70871	probable membrane pro 1.24e-02
3	88	18.7	607	2	F69769	conserved hypothetical 1.80e-01
4	84	17.8	362	2	S36766	prostaglandin E recep 6.87e-01
5	84	17.8	395	2	S36765	prostaglandin E recep 6.87e-01
6	84	17.8	387	2	S36767	prostaglandin E recep 6.87e-01
7	84	17.8	417	2	S36764	prostaglandin E recep 6.87e-01
8	82	17.4	198	2	S55599	superoxide dismutase 1.32e+00
9	80	17.0	373	2	A69158	sensory transduction 2.53e+00
10	80	17.0	895	2	F75894	hypothetical protein 2.53e+00
11	79	16.8	60	2	S59092	hypothetical protein 3.48e+00
12	79	16.8	615	1	ABCS	serum albumin precurs 3.48e+00
13	78	16.6	361	2	A53216	prostaglandin E2 rece 4.77e+00
14	78	16.6	363	2	B64807	abrB protein - Escher 4.77e+00
15	78	16.6	365	2	S51315	prostaglandin E recep 4.77e+00
16	78	16.6	385	2	I38748	prostaglandin receptor 4.77e+00
17	78	16.6	367	2	JC2056	prostaglandin E2 rece 4.77e+00
18	78	16.6	374	2	I38747	prostaglandin receptor 4.77e+00
19	78	16.6	374	2	S51317	prostaglandin E recep 4.77e+00
20	78	16.6	388	2	S51316	prostaglandin E recep 4.77e+00
21	78	16.6	388	2	I38750	prostaglandin receptor 4.77e+00
22	78	16.6	390	2	S51313	prostaglandin E recep 4.77e+00
23	78	16.6	390	2	S43375	prostaglandin E recep 4.77e+00

#journal Nature (1998) 393:537-544
#title Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.

#cross-references GB:AL021184; GB:AL123456; NID:g3261498; PID:e1237450;
#accession F70871

##status preliminary; nucleic acid sequence not shown;
##molecule_type DNA
##residues 1-591 ##label COL
##cross-references GB:AL021184; GB:AL123456; NID:g3261498; PID:e1237450;
#experimental_source strain H37Rv

GENETICS
#gene Rv1459c
#length 591 #molecular-weight 62693 #checksum 9361

Query Match 20.08; Score 94; DB 2; Length 591;
Best Local Similarity 39.08; Pred. No. 2.24e-02;
Matches 16; Conservative 6; Mismatches 18; Indels 1; Gaps 1;

Db 70 LMLPSRI-QTVSLTMTTGTGAVMMALAWMLGRFTLGRMR 109
QY 17 LLILOTRKPHGTGHLKTSGGMTVMVKKLLMTFMACRCGM 57

RESULT 3
ENTRY conserved hypothetical protein ydao - Bacillus subtilis
TITLE #formal_name Bacillus subtilis
ORGANISM 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
DATE 24-Sep-1998

ACCESSIONS F69769
REFERENCE A69580
#authors Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.;

Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.F.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, J.; Devine, K.M.; Duesterhoeft, A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.; Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glasner, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Giuseppe, G.; Guy, B.J.; Haga, K.; Halech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Kiehr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.; Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Pressac, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.; Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serrro, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Takakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenecker, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.; Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

#journal Nature (1997) 390:249-256
#title The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
#cross-references NID:98044033
#accession F69769

##status preliminary; nucleic acid sequence not shown;
##molecule_type DNA
##residues 1-607 ##label KUN
##cross-references GB:Z99106; GB:AL009126; NID:g2632653; PID:e1182398;
#experimental_source strain 168

GENETICS
#gene ydao
#length 607 #molecular-weight 66771 #checksum 4065

Query Match 18.7%; Score 88; DB 2; Length 607;
Best Local Similarity 32.7%; Pred. No. 1.80e-01;
Matches 16; Conservative 10; Mismatches 23; Indels 0; Gaps 0;

Db 181 LLVIAVGLFKLMTGIDQPAHHTSLGTPVAGITLFLLLKAFSSGCSAL 229
QY 9 LGVYKAGLLLIQTRKPHGTGHLKTSGGMTVMVKKLLMTFMACRCGM 57

RESULT 4
ENTRY #type complete
TITLE prostaglandin E receptor 3C - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change
29-Jan-1999

ACCESSIONS S36766
REFERENCE S36764
#authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;

Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S. Nature (1993) 365:166-170
#journal Alternative splicing of C-terminal tail of prostaglandin E receptor subtype EP3 determines G-protein specificity.
#cross-references MUID:93382504
#accession S36766
##status preliminary
##molecule_type mRNA
##residues 1-362 ##label NAM
#cross-references GB:D21347; NID:g456444; PID:d1005346; PID:g456445
CLASSIFICATION #superfamily prostaglandin E receptor EPI
SUMMARY #length 362 #molecular-weight 39887 #checksum 8225

Query Match 17.8%; Score 84; DB 2; Length 362;
Best Local Similarity 37.5%; Pred. No. 6.87e-01;
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWSSHMKTSVTRVLLGVWLAVLAF 187
QY 20 LQTRKPH-TGNHLKTSGGMTVMVKKLLMTF 50

RESULT 5
ENTRY #type complete
TITLE prostaglandin E receptor 3B - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change
29-Jan-1999

ACCESSIONS S36765
REFERENCE S36764
#authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;

Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S. Nature (1993) 365:166-170
#journal Alternative splicing of C-terminal tail of prostaglandin E receptor subtype EP3 determines G-protein specificity.
#cross-references MUID:93382504
#accession S36765
##status preliminary
##molecule_type mRNA
##residues 1-385 ##label NAM
#cross-references GB:D21346; NID:g415725; PID:d1005345; PID:g415726
CLASSIFICATION #superfamily prostaglandin E receptor EPI
SUMMARY #length 385 #molecular-weight 42466 #checksum 6243

```

Query Match      17.8%; Score 84; DB 2; Length 385;
Best Local Similarity 37.5%; Pred. No. 6.87e-01;
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWYSSHMKTSTVTRVLLGVWLVAVLAF 187
QY 20 LQTRKPH-TGNHLATSGGWTVMVKLLMTF 50

RESULT 6
ENTRY #type complete
TITLE prostaglandin E receptor - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 09-Dec-1993 #sequence_revision 27-Feb-1997 #text_change
ACCESSIONS S36767
REFERENCE S36764
#authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;
#journal Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S.
#title Nature (1993) 365:166-170
#description Alternative splicing of C-terminal tail of prostaglandin E
#cross-references MUID:93382504
#accession S36767
#status preliminary
#molecule_type mRNA
#residues 1-387 #label NAM
#cross-references GB:D21348; NID:g415729; PID:d1005347; PID:g415730
CLASSIFICATION #superfamily prostaglandin E receptor EPI
SUMMARY #length 387 #molecular-weight 42470 #checksum 1981

Query Match      17.8%; Score 84; DB 2; Length 387;
Best Local Similarity 37.5%; Pred. No. 6.87e-01;
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWYSSHMKTSTVTRVLLGVWLVAVLAF 187
QY 20 LQTRKPH-TGNHLATSGGWTVMVKLLMTF 50

RESULT 7
ENTRY #type complete
TITLE prostaglandin E receptor 3A - bovine
ORGANISM #formal_name Bos primigenius taurus #common_name cattle
DATE 10-Dec-1993 #sequence_revision 27-Feb-1997 #text_change
ACCESSIONS S36764
REFERENCE S36764
#authors Namba, T.; Sugimoto, Y.; Negishi, M.; Irie, A.; Ushikubi, F.;
#journal Kakizuka, A.; Ito, S.; Ichikawa, A.; Narumiya, S.
#title Nature (1993) 365:166-170
#description Alternative splicing of C-terminal tail of prostaglandin E
#cross-references MUID:93382504
#accession S36764
#status preliminary
#molecule_type mRNA
#residues 1-417 #label NAM
#cross-references GB:D21345; NID:g415723; PID:d1005344; PID:g415724
CLASSIFICATION #superfamily prostaglandin E receptor EPI
SUMMARY #length 417 #molecular-weight 46362 #checksum 9556

Query Match      17.8%; Score 84; DB 2; Length 417;
Best Local Similarity 37.5%; Pred. No. 6.87e-01;
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWYSSHMKTSTVTRVLLGVWLVAVLAF 187
QY 20 LQTRKPH-TGNHLATSGGWTVMVKLLMTF 50

RESULT 8
ENTRY #type complete

```

```

TITLE superoxide dismutase (EC 1.15.1.1) (Fe) - Plasmodium
ORGANISM falciparum
DATE 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change
ACCESSIONS S55499; S51750
REFERENCE S55499
#authors Becuwe, P.; Gratepanche, S.; Fourmaux, M.; van Beeumen, J.;
#journal Samyn, B.; Mercereau-Puijalon, O.; Touzel, J.; Slomianny,
#title C.; Camus, D.; Dive, D.
#description submitted to the EMBL Data Library, June 1995
#description Characterization of iron-dependent endogenous superoxide
#accession S55499
#molecule_type mRNA
#residues 1-198 #label BEC
#cross-references EMBL:Z49819; NID:g861076; PID:e183284; PID:g861077
REFERENCE S51750
#authors Becuwe, P.; Fourmaux, M.; Van Beeumen, J.; Samyn, B.;
#journal Slomianny, C.; Touzel, J.; Camus, D.; Dive, D.
#title submitted to the EMBL Data Library, September 1993
#description Characterization of iron-dependent endogenous SOD of
#accession S51750
#molecule_type DNA
#residues 32-160 #label BE2
#cross-references EMBL:Z25887; NID:g607155; PID:e85185; PID:g607156
FUNCTION catalyzes the dismutation of 2 molecules of peroxide radical
#description to dioxygen and hydrogen peroxide
CLASSIFICATION #superfamily superoxide dismutase (Mn)
KEYWORDS metalloprotein; oxidoreductase
FEATURE 27,74,158,152 #binding_site iron (His, His, Asp, His) #status
SUMMARY #length 198 #molecular-weight 22734 #checksum 2429
Query Match      17.4%; Score 82; DB 2; Length 198;
Best Local Similarity 34.1%; Pred. No. 1.32e-00;
Matches 15; Conservative 11; Mismatches 15; Indels 3; Gaps 2;

Db 119 TSGGWSWALNNNNKLVLOTH--DAGNPKDNTGPIILTCDIW 160
QY 2 YTSLSGHLGVKANGLLILQTRKFTGHNHKTSGGMVTWV-RKW 44

RESULT 9
ENTRY #type complete
TITLE sensory transduction histidine kinase - Methanobacterium
ORGANISM #formal_name Methanobacterium thermoautotrophicum
DATE 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change
ACCESSIONS A69158
REFERENCE A69000
#authors Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.;
#journal Dubois, J.; Aldredge, T.; Bashirzadeh, R.; Blakely, D.;
#title Cook, R.; Gilbert, R.; Harrison, D.; Hoang, L.; Keagle, P.;
#journal Wang, W.; Pothier, B.; Qiu, D.; Spadafora, R.; Vicaire, R.;
#title Lumm, Y.; Wierzbowski, J.; Gibson, R.; Jiawani, N.; Caruso,
#journal A.; Bush, D.; Sifer, H.; Patwell, D.; Prabhakar, S.;
#journal McDougall, S.; Shimer, G.; Goyal, A.; Pietrowski, S.;
#journal Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling,
#journal J.; Reeve, J.N.
#title J. Bacteriol. (1997) 179:7135-7155
#description Complete genome sequence of Methanobacterium
#cross-references MUID:98037314
#accession A69158
#status preliminary; nucleic acid sequence not shown;
#molecule_type DNA translation not shown

```

```

#accession S59092
##status. 1-nucleic acid sequence not shown; translation not shown
##molecule_type DNA
##residues 1-60 #label LEB
##cross-references EMBL:Z47547; NID:G1019057; PID:G1019071
##experimental_source female gametophytes
##note the nucleotide sequence was submitted to the EMBL Data
Library, January 1995

GENETICS
#genome mitochondrion
#genetic_code SGC3
KEYWORDS
SUMMARY #length 60 #molecular-weight 6633 #checksum 6517

Query Match 16.8%; Score 79; DB 2; Length 60;
Best Local Similarity 42.9%; Pred. No. 3.48e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0

Db 1 MVRHWLLMLAFKGS 14
||: |||::|:::
QY 40 MVRKWLLETFMAG 53

RESULT 12
ENTRY ABCBS #type complete
TITLE serum albumin precursor - chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change
05-Sep-1997
ACCESSIONS SI5571; A05078; A13451
REFERENCE SI5571
#authors Cassidy, A.I.; Salkild, C.K.; Baverstock, P.; Wallace, J.C.
#submission submitted to the EMBL Data Library, July 1991
#accession SI5571
##molecule_type mRNA
##residues 1-615 #label CAS
##cross-references EMBL:X60688; NID:G63747; PID:G63748
REFERENCE A05078
#authors Hache, R.J.G.; Wiskocil, R.; Vasa, M.; Roy, R.N.; Lau,
P.C.K.; Deeley, R.G.
#journal J. Biol. Chem. (1983) 258:4556-4564
#title The 5' noncoding and flanking regions of the avian very low
density apolipoprotein II and serum albumin genes.
Homologies with the egg white protein genes.
#cross-references MUID:83161037
#accession A05078
##molecule_type DNA
##residues 1-28 #label HAC
##cross-references GB:V00381; NID:G63038; PID:G63039
REFERENCE A13451
#authors Rosen, A.M.; Geller, D.M.
#journal Biochem. Biophys. Res. Commun. (1977) 78:1060-1066
#title Chicken microsomal albumin: amino terminal sequence of
chicken proalbumin.
#cross-references MUID:78019943
#accession A13451
##molecule_type protein
##residues 19-23, M', 25-30 #label ROS
COMMENT Serum albumin is synthesized in the liver as preproalbumin. It
binds copper, nickel, calcium (weakly, at 2-3 sites), and
protoporphyrin, long-chain fatty acids, prostaglandins, steroid
hormones (weak bonds with these hormones promote their transfer
across the membranes), thyroxine, and triiodothyronine.
CLASSIFICATION #superfamily serum albumin; serum albumin repeat homology
carrier protein; duplication; metal binding; plasma
FEATURES
1-18 #domain signal sequence #status predicted #label SIG
19-26 #domain propeptide #status predicted #label PRO
27-613 #product serum albumin #status predicted #label MAT
27-613 #domain serum albumin repeat homology #label SAI
225-398 #domain serum albumin repeat homology #label SA2
417-596 #domain serum albumin repeat homology #label SA3
30 #binding site copper (His) #status predicted

```

```

80-89,102-118,
117-128,152-197,
196-205,228-274,
273-281,293-307,
306-317,344-389,
388-397,420-466,
465-476,489-505,
504-515,542-587,
586-595
SUMMARY      #disulfide_bonds #status predicted
              #length 615 #molecular-weight 69918 #checksum 4822
Query Match      16.8%; Score 79; DB 1; Length 615;
Best Local Similarity 46.7%; Pred. No. 3.48e+00;
Matches 14; Conservative 6; Mismatches 7; Indels 3; Gaps 3;

Db 556 LLINIKRPQTEQIKTIADGFTAMVDK 585
QY 17 LLI-LQTRKPH-TGNHLKT-SGGWMTVMVK 43

RESULT 13
ENTRY      A53216      #type complete
TITLE      prostaglandin E2 receptor, subtype EP3 (clone 72A) - rabbit
ALTERNATE_NAMES  #formal_name Oryctolagus cuniculus #common_name domestic
ORGANISM    rabbit
DATE        19-Oct-1995 #sequence_revision 19-Oct-1995 #text_change
ACCESSIONS  A53216
REFERENCE    Breyer, R.M.; Emeson, R.B.; Tarrag, J.L.; Breyer, M.D.; Davis,
              L.S.; Abramson, R.M.; Ferrenbach, S.M.
              J. Biol. Chem. (1994) 269:6163-6169
#journal    J. Biol. Chem. (1994) 269:6163-6169
#title      Alternative splicing generates multiple isoforms of a rabbit
              prostaglandin E-2 receptor.
#cross-references MUID:94164982
#accession  A53216
#status     preliminary
#molecule_type mRNA
#residues   1-361 #label BRE
#cross-references GB:U04273; NID:q467978; PID:q467979
CLASSIFICATION #superfamily prostaglandin E receptor Epl
KEYWORDS      alternative splicing; transmembrane protein
SUMMARY      #length 361 #molecular-weight 40011 #checksum 4634

Query Match      16.6%; Score 78; DB 2; Length 361;
Best Local Similarity 31.3%; Pred. No. 4.77e+00;
Matches 10; Conservative 9; Mismatches 1; Indels 1; Gaps 1;

Db 153 LAIRAPHYASHMKTRATRAVLLGWLAVLAF 184
QY 20 LQTRKPH-TGNHLKTSGGWMTVMVKWLLMTF 50

RESULT 14
ENTRY      B64807      #type complete
TITLE      abrB protein - Escherichia coli
ALTERNATE_NAMES  #formal_name Escherichia coli
ORGANISM    Escherichia coli
DATE        12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change
ACCESSIONS  B64807
REFERENCE    Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.;
              Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.;
              Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.;
              Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao,
              Y.
              Science (1997) 277:1453-1462
#journal    Science (1997) 277:1453-1462
#title      The complete genome sequence of Escherichia coli K-12.
#cross-references MUID:97426617
#accession  B64807
#status     nucleic acid sequence not shown; translation not shown
#molecule_type DNA
#residues   1-363 #label BLAT

```

```

##cross-references GB:AE000174; GB:U00096; NID:q1786920; PID:q1786933;
##experimental_source strain K-12, substrain MG1655
GENETICS
#gene      abrB
#keywords   transmembrane protein
FEATURE
22-38      #domain transmembrane #status predicted #label TM1\
40-56      #domain transmembrane #status predicted #label TM2\
98-114     #domain transmembrane #status predicted #label TM3\
162-178    #domain transmembrane #status predicted #label TM4\
198-214    #domain transmembrane #status predicted #label TM5\
277-293    #domain transmembrane #status predicted #label TM6\
329-345    #domain transmembrane #status predicted #label TM7\
SUMMARY    #length 363 #molecular-weight 38939 #checksum 8760

Query Match      16.6%; Score 78; DB 2; Length 363;
Best Local Similarity 40.7%; Pred. No. 4.77e+00;
Matches 11; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

Db 228 AGAVLQ-SGLITIELPEWLLAMAYMA 253
QY 27 TGNHLKTSGGWMTM-VKKWLLMTFMA 52

RESULT 15
ENTRY      S51315      #type complete
TITLE      prostaglandin E receptor, subtype EP3 splice form b - human
ALTERNATE_NAMES  EP(3-III) prostanoind receptor
ORGANISM    #formal_name Homo sapiens #common_name man
DATE        07-May-1995 #sequence_revision 01-Sep-1995 #text_change
              17-Mar-1999
ACCESSIONS  S68995; S41078; S51315
REFERENCE    Schmid, A.; Thierauch, K.H.; Schleuning, W.D.; Dinter, H.
              Eur. J. Biochem. (1995) 228:23-30
#journal    Eur. J. Biochem. (1995) 228:23-30
#title      Splice variants of the human EP(3) receptor for prostaglandin
              E(2)
#cross-references MUID:95188908
#accession  S68995
#molecule_type mRNA
#residues     1-365 #label SC2
#cross-references EMBL:X83859; NID:q633211; PID:q633212
#note        Submitted to the EMBL Data Library, January 1995
REFERENCE    S41076
#authors     Adam, M.; Boie, Y.; Rushmore, T.H.; Mueller, G.; Bastien, L.;
              McKee, K.T.; Metters, K.M.; Abramovitz, M.
              FEBS Lett. (1994) 338:170-174
#journal    FEBS Lett. (1994) 338:170-174
#title      Cloning and expression of three isoforms of the human EP(3)
              prostanoid receptor.
#cross-references MUID:94139918
#accession  S41073
#status     preliminary
#molecule_type mRNA
#residues     1-365 #label ADA
GENETICS
#gene      GDB:PTGER3
#cross-references GDB:321928; OMIM:176806
#map_position lp31.2-lp31.2
CLASSIFICATION #superfamily prostaglandin E receptor EP1
KEYWORDS      alternative splicing
SUMMARY      #length 365 #molecular-weight 40501 #checksum 2560

Query Match      16.8%; Score 78; DB 2; Length 365;
Best Local Similarity 31.3%; Pred. No. 4.77e+00;
Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

Db 157 LAIRAFHWYASHMKTRATRAVLLGWLAVLAF 188
QY 20 LQTRKPH-TGNHLKTSGGWMTVMVKWLLMTF 50

Search completed: Wed Dec 22 10:26:09 1999

```

Tue Jan 4 11:52:17 2000

US-09-209-961-20.rpr

Page 6

Job time : 9 secs.

 W P E R E H

 (TM)

Release 3.1a John F. Collins, Biocomputing Research Unit.
 Copyright (c) 1993-1998 University of Edinburgh, U.K.
 Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
 Run on: Wed Dec 22 10:26:27 1999; MasPar time 4.57 Seconds
 Tabular output not generated. 364.668 Million cell updates/sec

Title: >US-09-209-961-20
 Description: (1-59) from US09209961.pep
 Perfect Score: 471
 Sequence: 1 MYTSLWGLGVKANGLLIL.....MVKKLLILMTFMAGCRGMIY 59

Scoring table: PAM 150
 Gap 11

Searched: 77977 seqs, 28268293 residues

Post-processing: Minimum Match 0%
 Listing first 45 summaries

Database: swiss-prot37
 1:swissprot

Statistics: Mean 37.670; Variance 59.374; scale 0.634

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	84	17.8	417	1 PER3_BOVIN	PROSTAGLANDIN E2 RECP	1.76e-01
2	83	17.6	373	1 PER3_PIG	PROSTAGLANDIN E2 RECP	2.58e-01
3	81	17.2	508	1 NO60_DROME	NUCLEOLAR PROTEIN AT B	5.28e-01
4	79	16.8	615	1 ABRU_CHICK	SERUM ALBUMIN PRECURSO	1.08e+00
5	78	16.6	363	1 ABRB_ECOLI	ABRB PROTEIN	1.53e+00
6	78	16.6	390	1 PER3_HUMAN	PROSTAGLANDIN E2 RECP	1.53e+00
7	78	16.6	411	1 PER3_RABIT	PROSTAGLANDIN E2 RECP	1.53e+00
8	78	16.6	547	1 MERA_STAUB	MERCURIC REDUCTASE (EC	1.53e+00
9	78	16.6	672	1 BGAL_BACST	BETA-GALACTOSIDASE I (1.53e+00
10	76	16.1	415	1 AMSJ_ERWAM	AMYLOVORAN BIOSYNTHESI	3.07e+00
11	76	16.1	564	1 AGGL_RICCO	AGGLUTININ PRECURSOR (3.07e+00
12	76	16.1	811	1 PRTA_RHORI	PRIMOSOMAL PROTEIN N'	3.07e+00
13	75	15.9	93	1 YSL1_THEFL	HYPOTHETICAL PROTEIN I	4.32e+00
14	75	15.9	152	1 UL04_HCMVA	EARLY GLYCOPROTEIN GP4	4.32e+00
15	75	15.9	398	1 Y05P_MYCTU	HYPOTHETICAL 43.7 KD P	4.32e+00
16	75	15.9	783	1 FUR1_XENLA	FURIN 1 PRECURSOR (EC	4.32e+00
17	75	15.9	1179	1 VG56_HSV1	HYPOTHETICAL GENE 56 P	4.32e+00
18	74	15.7	209	1 YC02_KLEPN	3-HYDROXYANTHRANILATE	6.06e+00
19	74	15.7	286	1 XHAO_HUMAN	HYPOPHOSPHATIDASE 2	6.06e+00
20	74	15.7	317	1 APE_PAPAN	APOLIPOPROTEIN E PRECU	6.06e+00
21	74	15.7	317	1 APE_MACFA	APOLIPOPROTEIN E PRECU	6.06e+00
22	74	15.7	317	1 APE_HUMAN	APOLIPOPROTEIN E PRECU	6.06e+00
23	74	15.7	519	1 NIFL_AZOVI	NITROGEN FIXATION REGU	6.06e+00

24 14 15.7 576 1 RIC1_RICCO
 25 74 15.7 773 1 CPT1_HUMAN
 26 74 15.7 971 1 SN2L_CAEEL
 27 73 15.5 102 1 DAP1_HUMAN
 28 73 15.5 104 1 VPR_HV2ST
 29 73 15.5 180 1 RBS_MUSAC
 30 73 15.5 434 1 ENO_LOLPE
 31 73 15.5 693 1 YET1_SCHPO
 32 73 15.5 1133 1 SRE1_CRIGR
 33 72 15.3 135 1 YDQ_ECOLI
 34 72 15.3 269 1 CDSA_BACSU
 35 72 15.3 399 1 FDHB_METFO
 36 72 15.3 584 1 PME_BRANA
 37 72 15.3 590 1 SYD_ECOLI
 38 72 15.3 618 1 Y4RQ_RHISN
 39 72 15.3 645 1 COX1_PARTE
 40 71 15.1 237 1 FGF3_XENLA
 41 71 15.1 529 1 ACH2_HUMAN
 42 71 15.1 1027 1 YDIJ_HAEIN
 43 71 15.1 1227 1 B3A3_RAT
 44 71 15.1 1515 1 GDE_HUMAN
 45 71 15.1 2594 1 7LES_DROVI

ALIGNMENTS

RESULT 1

ID PER3_BOVIN STANDARD; PRT; 417 AA.
 AC P34979:
 DT 01-FEB-1994 (REL. 28, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE
 DE RECEPTOR, EP3 SUBTYPE)
 GN PTGER3.
 OS BOS TAURUS (BOVIN)
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
 [1]
 SP SEQUENCE FROM N.A.
 KC TISSUE-ADRENAL MEDULLA;
 IX MEDLINE; 53382504.
 RA NAKAZUKA A., ITO S., ICHIKAWA A., NARUMIYA S.S.;
 RA "Alternat.ve splicing of C-terminal tail of prostaglandin E receptor
 RT subtype EP3 determines G-protein specificity";
 RL NATURE 365:166-170(1993).
 CC -!- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR
 CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION.
 CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL
 CC NEURONS. INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY
 CC TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF
 CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE
 CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF
 CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS INTERACT WITH DIFFERENT
 CC SECOND MESSENGER SYSTEMS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS (A TO D) THAT DIFFERS
 CC AT THE C-TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS
 CC RECEPTOR GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT
 CC DIFFERENT COUPLING PROPERTIES WITH G PROTEINS: EP3A COUPLES TO
 CC G(1)/G(0) PROTEINS; EP3B AND EP3C COUPLE TO G(S), AND EP3D COUPLES
 CC TO G(1), G(S) AND G(P).
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).

```

DR EMBL; D21345; G415724; -
DR EMBL; D21346; G415726; -
DR EMBL; D21347; G456445; -
DR EMBL; D21348; G415730; -
DR PIR; S36764; S36764.
DR GCRDB; GCR_0784; -
DR GCRDB; GCR_0785; -
DR GCRDB; GCR_0786; -
DR GCRDB; GCR_0787; -
DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
DR PFAM; PF00001; 7tm_1; 1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
FT DOMAIN 1 52 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 53 77 1 (POTENTIAL).
FT DOMAIN 78 90 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 91 111 2 (POTENTIAL).
FT DOMAIN 112 130 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 131 152 3 (POTENTIAL).
FT DOMAIN 153 173 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 175 196 4 (POTENTIAL).
FT DOMAIN 197 226 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 227 252 5 (POTENTIAL).
FT DOMAIN 253 282 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 283 306 6 (POTENTIAL).
FT DOMAIN 307 326 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 327 348 7 (POTENTIAL).
FT DOMAIN 349 417 CYTOPLASMIC (POTENTIAL).
FT VARSPLIC 359 387 LLKHSYGLDTEGTEKNDKEMENLYS ->
FT VARSPLIC 388 417 MISSING (IN EP3B).
FT VARSPLIC 359 362 LLKG -> HVGS (IN EP3C).
FT VARSPLIC 363 417 MISSING (IN EP3C).
FT VARSPLIC 359 385 ASPRSMWDPSPTRDRTRVPCIGSTES (IN EP3D).
FT VARSPLIC 386 417 MISSING (IN EP3D).
FT CARBOHYD 17 17 POTENTIAL.
FT CARBOHYD 35 35 POTENTIAL.
FT CARBOHYD 216 216 POTENTIAL.
FT CARBOHYD 307 307 POTENTIAL.
SQ SEQUENCE 417 AA; 46362 MW; 5AB01C72 CRC32;

Query Match 17.8%; Score 84; DB 1; Length 417;
Best Local Similarity 37.5%; Pred. No. 1.76e-01;
Matches 12; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

Db 156 LATRAPHWYSSHKMTSVTRAVLLGVWLVAVLAF 187
QY 20 LQTRKPH-TGNHLKTSKGMVTWKVLLMTF 50
| | | | | : : : : : | : : : :
| | | | | : : : : : | : : : :

RESULT 2
ID PER3_PIG STANDARD; PRT; 373 AA.
AC P50131; Q28967.
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE
DE RECEPTOR, EP3 SUBTYPE).
GN PIGR3
OS SUS SCROFA (PIG).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; SUIFORMES; SUINA; SUIDAE; SUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HEART.
RA MEYER J., SCHROER K.;
RL SUBMITTED (MAY-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
[2]
RP SEQUENCE OF 323-365 FROM N.A. (SHORT FORM).
RC TISSUE-HEART.
RA MEYER J., SCHROER K.;
RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DBJ DATA BANKS.

-1- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR
MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,
MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL
NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY
TUBULUS AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF
THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE
CYCLASE MEDIATED BY G-1 PROTEINS, AND TO AN ELEVATION OF
INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH
DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).
-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
-1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.

THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; U27083; G984358; -
EMBL; U30374; G984362; -
GCRDB; GCR_1576; -
PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
PFAM; PF00001; 7tm_1; 1.
KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
FT DOMAIN 1 30 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 31 55 1 (POTENTIAL).
FT DOMAIN 56 68 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 69 89 2 (POTENTIAL).
FT DOMAIN 90 108 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 109 130 3 (POTENTIAL).
FT DOMAIN 131 151 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 153 174 4 (POTENTIAL).
FT DOMAIN 175 204 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 205 230 5 (POTENTIAL).
FT DOMAIN 231 260 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 261 284 6 (POTENTIAL).
FT DOMAIN 285 304 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 305 326 7 (POTENTIAL).
FT DOMAIN 327 373 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 13 13 POTENTIAL.
FT CARBOHYD 194 194 POTENTIAL.
FT DISULFID 107 185 BY SIMILARITY.
FT VARSPLIC 337 373 AVSOKQREAAALFTLHLSISRTPEGEARVLFKSKC ->
VANAVCSCKNGQKQVQISLSHSITQTEA (IN SHORT
FORM).
SQ SEQUENCE 373 AA; 41416 MW; BFECCOD9 CRC32;

Query Match 17.6%; Score 83; DB 1; Length 373;
Best Local Similarity 34.4%; Pred. No. 2.55e-01;
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Db 134 LAIRAPHWYSSHKMTSATRAVLLGVWLVAVLAF 165
QY 20 LQTRKPH-TGNHLKTSKGMVTWKVLLMTF 50
| | | | | : : : : : | : : : :
| | | | | : : : : : | : : : :

RESULT 3
ID NO60_DROME STANDARD; PRT; 508 AA.
AC O44081;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE NUCLEOLAR PROTEIN AT BAND 60B.
GN NOP60B.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.

```

RA POOLE S.J., BILLIN A., CADWELL C., ERICKSON C., BUCHHOLZ R.,
 RA MERRIAM J., CARBON J., TO EMBL/GENBANK/DBJ DATA BANKS.
 RL SUBMITTED (AUG-1997) TO EMBL/GENBANK/DBJ DATA BANKS.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR.
 CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF017230; G2661227;
 DR FLYBASE; FBgn0023184; Nop60B.
 KW NUCLEAR PROTEIN.
 FT DOMAIN 8 16 POLY-LYS.
 FT DOMAIN 450 453 POLY-ALA.
 FT DOMAIN 461 471 POLY-LYS.
 FT DOMAIN 489 499 POLY-LYS.
 SQ SEQUENCE 508 AA: 56830 MW; 09948431 CRC32;
 Query Match 17.2%; Score 81; DB 1; Length 508;
 Best Local Similarity 39.4%; Pred. No. 5.28e-01;
 Matches 13; Conservative 8; Mismatches 11; Indels 1; Gaps 1;
 Db 230 HLGVLGVGGOMLELRVRSGIQ-SERDGMVTM 261
 Qy 8 HLGVVKANGLLILOTRKPHGNHKTSGGVTM 40
 RESULT 4
 ID ALBUCHICK STANDARD; PRT; 615 AA.
 AC P19121;
 DT 01-NOV-1990 (REL. 16, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE SERUM ALBUMIN PRECURSOR.
 GN ALB.
 OS GALLUS GALLUS (CHICKEN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ARCHOSAURIA; AVES;
 OC NEOGNATHAE; GALLIFORMES; PHASIANIDAE; PHASIANINAE; GALLUS.
 RN
 RP SEQUENCE FROM N.A.
 RC TISSUE-LIVER;
 RA CASSADY A.I., SALKILD C.K., BAVERSTOCK P., WALLACE J.C.;
 RL SUBMITTED (JUL-1991) TO EMBL/GENBANK/DBJ DATA BANKS.
 RN
 RP SEQUENCE OF 1-28 FROM N.A.
 RX MEDLINE; 83161037.
 RA HACHE R.J.G., WISKOCIL R., VASA M., ROY R.N., LAU P.C.K., DEELEY R.G.;
 RT "The 5' noncoding and flanking regions of the avian very low density
 RT apolipoprotein II and serum albumin genes. Homologies with the egg
 RT white protein genes."
 RL J. BIOL. CHEM. 258:4556-4564(1983).
 RN
 RP SEQUENCE OF 19-30.
 RX MEDLINE; 78019943.
 RA ROSEN A.M., GELLER D.M.;
 RT "Chicken microsomal albumin: amino terminal sequence of chicken
 RT prealbumin."
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 78:1060-1066(1977).
 CC -1- FUNCTION: SERUM ALBUMIN, THE MAIN PROTEIN OF PLASMA, HAS A GOOD
 CC BINDING CAPACITY FOR WATER, CA⁺⁺, NA⁺, K⁺, FATTY ACIDS, HORMONES,
 CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE
 CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -1- TISSUE SPECIFICITY: PLASMA.
 CC -1- DOMAIN: COMPOSED OF THREE HOMOLOGOUS DOMAINS.
 CC -1- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X60688; G63748;
 DR EMBL; V00381; G63039;
 DR PIR; S15571; ABCHS.
 DR PROSITE; PS00212; ALBUMIN; 3.
 DR PFAM; PF00273; transport_prot; 1.
 DR HSSP; P02759; IUOR.
 KW PLASMA; METAL-BINDING; LIPID-BINDING; ALBUMIN; REPEAT; SIGNAL;
 KW COPPER.
 FT SIGNAL 1 18
 FT PROPEP 19 23
 FT CHAIN 24 615 SERUM ALBUMIN.
 FT REPEAT 31 206 1.
 FT REPEAT 225 398 2.
 FT REPEAT 417 596 3.
 FT METAL 30 30 COPPER (BY SIMILARITY).
 FT DISULFID 80 89 BY SIMILARITY.
 FT DISULFID 102 118 BY SIMILARITY.
 FT DISULFID 117 128 BY SIMILARITY.
 FT DISULFID 152 197 BY SIMILARITY.
 FT DISULFID 196 205 BY SIMILARITY.
 FT DISULFID 228 274 BY SIMILARITY.
 FT DISULFID 273 281 BY SIMILARITY.
 FT DISULFID 293 307 BY SIMILARITY.
 FT DISULFID 306 317 BY SIMILARITY.
 FT DISULFID 344 389 BY SIMILARITY.
 FT DISULFID 388 397 BY SIMILARITY.
 FT DISULFID 420 466 BY SIMILARITY.
 FT DISULFID 465 476 BY SIMILARITY.
 FT DISULFID 489 505 BY SIMILARITY.
 FT DISULFID 504 515 BY SIMILARITY.
 FT DISULFID 542 587 BY SIMILARITY.
 FT DISULFID 586 595 BY SIMILARITY.
 FT CARBOHYD 500 500 POTENTIAL.
 FT CONFLICT 24 24 F -> M (IN REF. 3).
 SQ SEQUENCE 615 AA; 69918 MW; DC5E6EEA CRC32;
 Query Match 16.8%; Score 79; DB 1; Length 615;
 Best Local Similarity 46.7%; Pred. No. 1.08e+00;
 Matches 14; Conservative 6; Mismatches 7; Indels 3; Gaps 3;
 Db 556 LLINIKRKPMTEEQIKTIADGFTAMVDK 585
 Qy 17 LLI-LQTRKPH-TGNHLKT-SGGWVTMVKK 43
 RESULT 5
 ID ABRB_ECOLI STANDARD; PRT; 363 AA.
 AC P75747;
 DT 01-NOV-1997 (REL. 35, CREATED)
 DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE ABRB PROTEIN.
 CN ABRB.
 OS ESCHERICHIA COLI.
 OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; ENTEROBACTERIACEAE;
 CC ESCHERICHIA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX STRAIN-K12 / MG1655;
 RX MEDLINE; 97426617.
 RA BLATNER F.R., PLUNKETT G. III, BLOCH C.A., PERNA N.T., BURLAND V.,
 RA RILEY M., COLLADO-VIDES J., GLASNER F.D., RODE C.K., MAYHEW G.F.,
 RA GREGOR J., DAVIS N.W., KIRKPATRICK H.A., GOEDEN M.A., ROSE D.J.,
 RA MAU B., SHAO Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL SCIENCE 277:1453-1474(1997).

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000174; GI786933;
 CC ECOGENE: EG13310; ABRB.
 CC
 CC TRANSMEMBRANE
 CC TRANSMEM 35 55 POTENTIAL.
 CC TRANSMEM 74 94 POTENTIAL.
 CC TRANSMEM 98 118 POTENTIAL.
 CC TRANSMEM 153 173 POTENTIAL.
 CC TRANSMEM 188 208 POTENTIAL.
 CC TRANSMEM 211 231 POTENTIAL.
 CC TRANSMEM 236 256 POTENTIAL.
 CC TRANSMEM 274 294 POTENTIAL.
 CC TRANSMEM 303 323 POTENTIAL.
 CC TRANSMEM 330 350 POTENTIAL.
 CC SEQUENCE 363 AA; 38939 MW; BA34DC91 CRC32;

Query Match 16.6%; Score 78; DB 1; Length 363;
 Best Local Similarity 40.7%; Pred. NO. 1.53e+00;
 Matches 11; Conservative 7; Mismatches 7; Indels 2; Gaps 2;

Db 228 AGAVLO-SGOLITELPEWLLAMAYMA 253

QY 27 TGNHLKTSGGWVTN-VKKWLLMTFMA 52

RESULT 6

ID PER3 HUMAN STANDARD; PRT; 390 AA.
 AC P43115; Q12943; Q12944; Q12945; O16546;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-JUL-1998 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE
 DE RECEPTOR, EP3 SUBTYPE).
 GN PTGER3.
 OS HOMO SAPIENS (HUMAN).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94139918.
 RA ADAM M., BOYE Y., RUSHMORE T.H., MULLER G., BASTIEN L., MCKEE K.T.,
 RA METTERS K.M., ABRAMOVITZ M.;
 RA "Cloning and expression of three isoforms of the human EP3 prostanoid
 RT receptor";
 RL FEBS LETT. 338:170-174(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-UTERUS;
 RX MEDLINE; 95188908.
 RA SCHMID A., THIERAUCH K.H., SCHLEUNING W.D., DINTER H.;
 RA "Splice variants of the human EP3 receptor for prostaglandin E2";
 RL EUR. J. BIOCHEM. 228:23-30(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE; 94161771.
 RA YANG J., XIA M., GOETZL E.J., AN S.;
 RT "Cloning and expression of the EP3-subtype of human receptors for
 RL prostaglandin E2";
 RN BIOCHEM. BIOPHYS. RES. COMMUN. 198:999-1006(1994).
 [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 94183149.
 RA KUNAPULI S.P., FEN MAO G., BASTEPE M., LIU-CHEN L.-Y., LI S.,

RA CHEUNG P.P., DERIEL J.K., ASHBY B.;
 RT "Cloning and expression of a prostaglandin E receptor EP3 subtype
 RL from human erythrocytes cells";
 RN BIOCHEM. J. 298:263-267(1994).
 [5]
 RP SEQUENCE FROM N.A.
 RC TISSUE-UTERUS;
 RX MEDLINE; 95072021.
 RA AN S., YANG J., SO S.W., ZENG L., GOETZL E.J.;
 RT "Isoforms of the EP3 subtype of human prostaglandin E2 receptor
 RL transduce both intracellular calcium and cAMP signals";
 RN BIOCHEMISTRY. 33:14496-14502(1994).
 [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE-KIDNEY;
 RX MEDLINE; 96074809.
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., MORI K., ICHIKAWA A.,
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;
 RT "Molecular cloning and expression of multiple isoforms of human
 RL prostaglandin E receptor EP3 subtype generated by alternative
 RN messenger RNA splicing: multiple second messenger systems and
 tissue-specific distributions";
 RL MOL. PHARMACOL. 48:869-879(1995).
 [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE-SMALL INTESTINE;
 RX MEDLINE; 94356288.
 RA REGAN J.W., BAILEY T.J., DONELLO J.E., PIERCE K.L., PEPPERL D.J.,
 RA ZHANG D., KEDZIE K.M., FAIRBAIRN C.E., BOGARDUS A.M., WOODWARD D.F.,
 RA GIL D.W.;
 RT "Molecular cloning and expression of human EP3 receptors: evidence of
 RL three variants with differing carboxyl termini";
 RN BR. J. PHARMACOL. 112:377-385(1994).
 [8]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 97230456.
 RA KOTANI M., TANAKA I., OGAWA Y., USUI T., TAMURA N., MORI K.,
 RA NARUMIYA S., YOSHIMI T., NAKAO K.;
 RT "Structural organization of the human prostaglandin EP3 receptor
 RL subtype gene (PTGER3)";
 RN GENOMICS 40:425-434(1997).
 CC -1- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2); THE EP3 RECEPTOR
 CC MAY BE INVOLVED IN INHIBITION OF GASTRIC ACID SECRETION,
 CC MODULATION OF NEUROTRANSMITTER RELEASE IN CENTRAL AND PERIPHERAL
 CC NEURONS, INHIBITION OF SODIUM AND WATER REABSORPTION IN KIDNEY
 CC TUBULES AND CONTRACTION IN UTERINE SMOOTH MUSCLE. THE ACTIVITY OF
 CC THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE
 CC CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF
 CC INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH
 CC DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT LEAST IN SMALL INTESTINE, HEART,
 CC AND PANCREAS.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST SIX FORMS THAT DIFFERS AT THE C-
 CC TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS RECEPTOR
 CC GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT
 CC COUPLING PROPERTIES WITH G PROTEINS. THE FORMS SHOWN HERE IS KNOWN
 CC AS EP3-I, EP3A, OR EP3AL.
 CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -1- CAUTION: IN REF.7 EP3B IS KNOWN AS EP3E, EP3C AS EP3D AND EP3D AS
 CC EP3F.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: S69200; G545304;
 CC EMBL: L27490; G440314;
 CC EMBL: L27488; G440310;

DR EMBL; L27489; G440312; -
 DR EMBL; X83857; G633208; -
 DR EMBL; X83858; G633210; -
 DR EMBL; X83859; G633212; -
 DR EMBL; X83860; G633214; -
 DR EMBL; X83861; G633216; -
 DR EMBL; X83862; G633218; -
 DR EMBL; X83863; G633220; -
 DR EMBL; L26976; G483352; -
 DR EMBL; S69326; G545919; -
 DR EMBL; L32660; G857668; -
 DR EMBL; L32661; G484162; -
 DR EMBL; L32662; G484164; -
 DR EMBL; D38297; G1199759; -
 DR EMBL; D38298; G1199765; -
 DR EMBL; D38299; G1199767; -
 DR EMBL; D38300; G1199761; -
 DR EMBL; D38301; G1199763; -
 DR EMBL; U13214; G532738; -
 DR EMBL; U13215; G532740; -
 DR EMBL; U13216; G532742; -
 DR EMBL; U13217; G532744; -
 DR EMBL; U13218; G532746; -
 DR EMBL; D86096; D1020741; -
 DR EMBL; D86087; D1020741; JOINED.
 DR EMBL; D86088; D1020741; JOINED.
 DR GCRDB; GCR_0837; -
 DR GCRDB; GCR_0886; -
 DR GCRDB; GCR_0887; -
 DR GCRDB; GCR_0888; -
 DR GCRDB; GCR_1279; -
 DR GCRDB; GCR_1288; -
 DR GCRDB; GCR_1809; -
 DR GCRDB; GCR_1810; -
 DR GCRDB; GCR_1811; -
 DR GCRDB; GCR_1812; -
 DR GCRDB; GCR_1813; -
 DR GCRDB; GCR_1866; -
 DR GCRDB; GCR_1867; -
 DR GCRDB; GCR_1868; -
 DR GCRDB; GCR_2037; -
 DR GCRDB; GCR_2038; -
 DR GCRDB; GCR_2039; -
 DR GCRDB; GCR_2041; -
 DR GCRDB; GCR_2099; -
 DR GCRDB; GCR_2100; -
 DR GCRDB; GCR_2101; -
 DR GCRDB; GCR_2102; -
 DR GCRDB; GCR_2103; -
 DR GCRDB; GCR_2104; -
 DR GCRDB; GCR_2105; -
 DR GCRDB; GCR_2660; -
 DR GCRDB; GCR_2661; -
 DR GCRDB; GCR_2662; -
 DR GCRDB; GCR_2663; -
 DR GCRDB; GCR_2664; -
 DR GCRDB; GCR_2665; -
 DR GCRDB; GCR_2666; -
 DR GCRDB; GCR_2667; -
 DR GCRDB; GCR_2668; -
 DR MIM; 176806; -
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
 DR PFAM; PF00001; 7cm_1; 1.
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW LIPOPROTEIN; PALMITATE; PHOSPHORYLATION; ALTERNATIVE SPLICING.
 FT DOMAIN 1 53
 FT TRANSMEM 54 78
 FT DOMAIN 79 91
 FT TRANSMEM 92 112
 FT DOMAIN 113 131
 FT TRANSMEM 132 153
 FT DOMAIN 154 174
 FT TRANSMEM 175 197
 FT TRANSMEM 176 197

HP DOMAIN 198 227
 FT TRANSMEM 228 253
 FT DOMAIN 254 283
 FT TRANSMEM 284 307
 FT DOMAIN 308 327
 FT TRANSMEM 328 349
 FT DOMAIN 350 390
 FT CARBOHYD 18 36
 FT CARBOHYD 36 36
 FT VARSPLIC 360 390
 FT VARSPLIC 360 390
 FT VARSPLIC 360 390
 FT VARSPLIC 360 390
 Note. remainder of annotations omitted.
 Query Match 16.6%; Score 78; DB 1; Length 390;
 Best Local Similarity 31.3%; Pred. No. 1.53e+00;
 Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;
 Db 157 LAIRAPHTASHMKTRATRAVLGVWLVAF 188
 QY 20 LOTRKPH-TGNHLKTSQGMVTVKWLMTF 50
 RESULT 7
 ID PER3.RABIT STANDARD; PRT; 411 AA.
 AC P46059;
 DT 01-NOV-1995 (REL. 32, CREATED)
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE PROSTAGLANDIN E2 RECEPTOR, EP3 SUBTYPE (PROSTANOID EP3 RECEPTOR) (PGE
 RECEPTOR, EP3 SUBTYPE)
 OS PTGER3.
 OS ORYCTOLAGUS CUNICULUS (RABBIT).
 CC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 CC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NEW ZEALAND WHITE; TISSUE=KIDNEY CORTEX;
 EN MEDLINE; 94164982.
 AA BIEVER R.M., FERRENBACH S.M.;
 RA ABROMSON R.M., FERRENBACH S.M.; BREYER M.D., DAVIS L.S.,
 XT "Alternative splicing generates multiple isoforms of a rabbit
 prostaglandin E2 receptor."
 RT J. BIOL. CHEM. 269:6163-6169(1994).
 CC -!- FUNCTION: RECEPTOR FOR PROSTAGLANDIN E2 (PGE2). THE ACTIVITY OF
 THIS RECEPTOR CAN COUPLE TO BOTH THE INHIBITION OF ADENYLATE
 CYCLASE MEDIATED BY G-I PROTEINS, AND TO AN ELEVATION OF
 INTRACELLULAR CALCIUM. THE VARIOUS FORMS CAN INTERACT WITH
 DIFFERENT SECOND MESSENGER SYSTEMS (BY SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -!- TISSUE SPECIFICITY: IN THE KIDNEY, HIGH LEVELS OF EXPRESSION ARE
 SEEN IN MEDULLARY THICK ASCENDING LIMB, AND LOWER LEVELS IN THE
 CORTICAL AND MEDULLARY COLLECTING DUCTS.
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST FOUR FORMS THAT DIFFERS AT THE C-
 TERMINUS ARE PRODUCED BY ALTERNATIVE SPLICING OF THIS RECEPTOR
 GENE. THEY HAVE IDENTICAL LIGAND BINDING PROPERTIES BUT DIFFERENT
 COUPLING PROPERTIES WITH G PROTEINS. THE SEQUENCE OF CLONE 77A IS
 SHOWN:
 CC -!- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib.ch).
 CC

DR EMBL; U04274; G467981; -
 DR EMBL; U04273; G467979; -
 DR EMBL; U04275; G467983; -
 DR EMBL; U04276; G467985; -
 DR GCRDB; GCR_0954; -
 DR GCRDB; GCR_0955; -
 DR PROSITE; PS00237; G_PROTEIN_RECEPTOR; FALSE_NEG.
 KW PFAM; PF00001; 7tm1; 1.
 KW G-PROTEIN COUPLED RECEPTOR; TRANSMEMBRANE; GLYCOPROTEIN;
 KW ALTERNATIVE SPLICING.
 FT DOMAIN 1 49
 FT TRANSMEM 50 74
 FT DOMAIN 75 87
 FT TRANSMEM 88 108
 FT DOMAIN 109 127
 FT TRANSMEM 128 149
 FT DOMAIN 150 170
 FT TRANSMEM 172 193
 FT DOMAIN 194 223
 FT TRANSMEM 224 249
 FT DOMAIN 250 279
 FT TRANSMEM 280 303
 FT DOMAIN 304 323
 FT TRANSMEM 324 345
 FT DOMAIN 346 411
 FT CARBOHYD 18 18
 FT CARBOHYD 32 32
 FT CARBOHYD 213 213
 FT DISULFID 126 204
 FT VARIANT 243 243
 FT VARSPIC 356 361
 FT VARSPIC 362 411
 FT VARSPIC 377 377
 FT VARSPIC 378 411
 FT VARSPIC 356 389
 FT VARSPIC 390 411
 SQ SEQUENCE 411 AA; 45600 MW; 299918C2 CRC32;
 Query Match 16.6%; Score 78; DB 1; Length 411;
 Best Local Similarity 31.3%; Pred. No. 1.53e+00;
 Matches 10; Conservative 9; Mismatches 12; Indels 1; Gaps 1;
 Db 153 LAIRAPHWYASHMKTRATRAVLGWLAVLAF 184
 QY 20 LQTRKPH-TGNHLKTSGMVTWVKKLLMTF 50
 RESULT 8
 ID MERA_STRAU STANDARD; PRT; 547 AA.
 AC P08663;
 DT 01-JAN-1988 (REL. 06, CREATED)
 DT 01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
 DT 01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)
 DE MERCURIC REDUCTASE (EC 1.16.1.1) (HG(II) REDUCTASE).
 GN MERA.
 OS STAPHYLOCOCCUS AUREUS.
 OG PLASMID P1258.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 CC STAPHYLOCOCCUS.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 87260937.
 RA LADDAGA R.A., CHU L., MISRA T.K., SILVER S.;
 RT "Nucleotide sequence and expression of the mercurial-resistance
 operon from Staphylococcus aureus plasmid p1258".
 RL PROC. NATL. ACAD. SCI. U.S.A. 84:5106-5110(1987).
 CC -!- FUNCTION: RESISTANCE TO HG(2+) IN BACTERIA APPEARS TO BE GOVERNED
 CC BY A SPECIALIZED SYSTEM WHICH INCLUDES MERCURIC REDUCTASE. MERA
 CC PROTEIN IS RESPONSIBLE FOR VOLATILIZING MERCURY AS HG(0).
 CC -!- CATALYTIC ACTIVITY: HG + NADP(+) + H(+) = HG(2+) + NADPH.
 CC -!- COFACTOR: FAD.
 CC -!- SUBUNIT: HOMODIMER.
 CC -!- THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE
 CC OXIDOREDUCTASES CLASS-I.
 CC -!- SIMILARITY: CONTAINS A COPY OF THE HEAVY-METAL-ASSOCIATED (HMA)
 CC DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L29436; G459907; -
 DR PIR; E29504; E29504
 DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
 DR PROSITE; PS01047; HMA; 1.
 DR PFAM; PF00070; pyr_redox; 1.
 DR PFAM; PF00403; HMA; 1.
 DR HSSP; P11959; 1EBD.
 KW MERCURIC RESISTANCE; OXIDOREDUCTASE; FLAVOPROTEIN; FAD; NADP;
 KW MERCURY; REDOX-ACTIVE CENTER; METAL-BINDING; PLASMIID.
 KW DOMAIN 10 38
 FT NP_BIND 87 117
 FT DISULFID 123 128
 FT NP_BIND 378 388
 FT METAL 544 544
 FT METAL 545 545
 SQ SEQUENCE 547 AA; 58565 MW; 60DAFC29 CRC32;
 Query Match 16.6%; Score 78; DB 1; Length 547;
 Best Local Similarity 26.7%; Pred. No. 1.53e+00;
 Matches 16; Conservative 14; Mismatches 28; Indels 2; Gaps 2;
 Db 326 IYIEVNGEQVIEADQVLVATGRKPTET-LNLESAGVTKGKGVLTNEYLOTNNRIY 384
 QY 1 MYTSLWGLGVVANGLLILOTRKPHNLKTSGMVTWVKKLLMT-FWAGCRGMIV 59
 RESULT 9
 ID BGAL_BACST STANDARD; PRT; 672 AA.
 AC P19668;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
 DE BETA-GALACTOSIDASE I (EC 3.2.1.23) (LACTASE).
 GN BGAB.
 OS BACILLUS STEAROTHERMOPHILUS.
 OC BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;
 CC BACILLUS.
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-14.
 RC STRAIN-ATCC 8005 / IAM11001;
 RC MEDLINE; 86223784.
 RA HIRATA H., FUKAZAWA T., NEGORO S., OKADA H.;
 RT "Structure of a beta-galactosidase gene of Bacillus
 stearothermophilus".
 RL J. BACTERIOL. 166:722-727(1986).
 CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL, NON-REDUCING BETA-D-
 CC GALACTOSE RESIDUES IN BETA-D-GALACTOSIDES.
 CC -!- BETA-GALACTOSIDASE I IS MORE THERMOSTABLE THAN THE II, AND THE
 CC III ISOFORMS, BEING STABLE UP TO 70 DEGREES CELSIUS.
 CC -!- SIMILARITY: BELONGS TO FAMILY 42 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----


```

RESULT 12
ID PRIA_RHORU STANDARD; PRT; 811 AA.
AC P05445;
DT 01-NOV-1988 (REL. 09, CREATED)
DT 01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE PRIMOSOMAL PROTEIN N' (REPLICATION FACTOR Y).
GN PRIA.
OS RHODOSPIRILLUM RUBRUM.
OC BACTERIA; PROTEOBACTERIA; ALPHA SUBDIVISION; RHODOSPIRILLACEAE;
OC RHODOSPIRILLUM.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 85251588.
RA FALK G., HAMPE A., WALKER J.E.;
RL BIOCHEM. J. 228:391-407(1985).
CC -1- FUNCTION: PRIA RECOGNIZES A SPECIFIC HAIRPIN SEQUENCE ON PHIX
CC SSDNA. THIS STRUCTURE IS THEN RECOGNIZED AND BOUND BY PROTEINS
CC PRIB AND PRIC. FORMATION OF THE PRIMOSOME PROCEEDS WITH THE
CC SUBSEQUENT ACTIONS OF DNAB, DNAC, DNAT AND PRIMASE. PRIA THEN
CC FUNCTIONS AS A HELICASE WITHIN THE PRIMOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PRIA SUBFAMILY OF HELICASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X02499; G46363;
DR EMBL; X02499; G46364; ALT_INIT.
DR PIR; S08579; S08579.
KW DNA REPLICATION; DNA-BINDING; ATP-BINDING; HELICASE; PRIMOSOME;
KW ZINC-FINGER.
FT NP_BIND 253 260 ATP (BY SIMILARITY).
FT SITE 349 352 DEEH BOX.
FT 2N_FING 520 532 C4-TYPE (POTENTIAL).
FT 2N_FING 547 563 C4-TYPE (POTENTIAL).
SQ SEQUENCE 811 AA; 85574 MW; 5F50C3D8 CRC32;

Query Match 16.1%; Score 76; DB 1; Length 811;
Best Local Similarity 34.7%; Pred. No. 3.07e+00;
Matches 17; Conservative 12; Mismatches 16; Indels 4; Gaps 4;

Db 754 LGVPAP-LAMLGRHRH-LLKAAAGVKVQPVVRHWSLVSTPPGVK 800
QY 9 LGVYKANGLLIOTKRPHTGNHLKTS-GGMV-TWVKWLLMTFMAGCR 55

RESULT 13
ID YSC1_THREFL STANDARD; PRT; 93 AA.
AC P25124;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL PROTEIN IN SC5B 5' REGION (ORFA) (FRAGMENT).
OS THERMUS AQUATICUS (SUBSP. FLAVUS).
OC BACTERIA; THERMUS/DEINOCOCCUS GROUP; THERMUS GROUP; THERMUS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=AT-62;
RC MEDLINE; 91238680.
RA NISHIYAMA M., HORINOCHI S., BEPPU T.;
RT "Characterization of an operon encoding succinyl-CoA synthetase and
RT malate dehydrogenase from Thermus flavus AT-62 and its expression in
RT Escherichia coli."
RL MOL. GEN. GENET. 226:1-9(1991).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X54073; G48175;
DR PIR; S15948; S15948.
KW HYPOTHETICAL PROTEIN.
FT NON_TER 1
SQ SEQUENCE 93 AA; 10011 MW; 0B029A19 CRC32;

Query Match 15.9%; Score 75; DB 1; Length 93;
Best Local Similarity 35.9%; Pred. No. 4.32e+00;
Matches 14; Conservative 8; Mismatches 14; Indels 3; Gaps 3;

Db 20 ALRPOLOVDTLKLILKEGPHOGVGYLE-DGSMVVV 57
QY 4 SLWGLGVKANGLLIL-QTRPHTG-NHLKTSGGNVTM 40

RESULT 14
ID UL04_HCMVA STANDARD; PRT; 152 AA.
AC P17146;
DT 01-AUG-1990 (REL. 15, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE EARLY GLYCOPROTEIN GP48 PRECURSOR.
GN UL4.
OS HUMAN CYTOMEGALOVIRUS (STRAIN AD169).
OC VIRUSES; DSDNA VIRUSES, NO RNA STAGE; HERPESVIRIDAE;
OC BETAHERPESVIRINAE; CYTOMEGALOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 90269039.
RA CHEE M.S., BANKIER A.T., BECK S., BOHNI R., BROWN C.M., CERNY R.,
RA HORSNELL T., HUTCHISON C.A. III, KOZARIDES T., MARTIGNETTI J.A.,
RA PREDDIE E., SATCHWELL S.C., TOMLINSON P., WESTON K.M., BARRELL B.G.;
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169."
RL CURR. TOP. MICROBIOL. IMMUNOL. 154:125-169(1990).
CC -1- FUNCTION: SINCE GP48 WAS FOUND IN THE VIRION, IT IS CONSIDERED
CC AN EARLY STRUCTURAL PROTEIN.
CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
CC -1- PTM: N-GLYCOSYLATED AND POSSIBLY O-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO RL11 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X17403; G59609;
DR PIR; S09767; VGBEY9.
KW EARLY PROTEIN; GLYCOPROTEIN; MEMBRANE; SIGNAL.
FT SIGNAL 1 25
FT CHAIN 26 152 EARLY GLYCOPROTEIN GP48.
FT CARBOHYD 48 48 POTENTIAL.
FT CARBOHYD 53 53 POTENTIAL.
FT CARBOHYD 61 61 POTENTIAL.
FT CARBOHYD 69 69 POTENTIAL.
FT CARBOHYD 108 108 POTENTIAL.
FT CARBOHYD 112 112 POTENTIAL.
FT CARBOHYD 122 122 POTENTIAL.
FT CARBOHYD 139 139 POTENTIAL.
FT CARBOHYD 148 148 POTENTIAL.
SQ SEQUENCE 152 AA; 17750 MW; 0BDE403 CRC32;

Query Match 15.9%; Score 75; DB 1; Length 152;

```

Best Local Similarity 45.0%; Pred. No. 4.32e+00;
Matches 9; Conservative 5; Mismatches 4; Indels 2; Gaps 2;

Db 1 MYMLRTWRLLPMVLLAAYC 20
QY 37 MYTWYKKW-LLLMTFMAG-C 54

RESULT 15
ID YOSP_MYCTU STANDARD; PRT; 398 AA.
AC Q10899;
DT 01-OCT-1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 43.7 KD PROTEIN CY251.25.
GN MTCY251.25.
OS MYCOBACTERIUM TUBERCULOSIS.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERINEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RA CONNOR R.; CHURCHER C.M.; BARRELL B.G.; RAJANDREAM M.A.; WALSH S.V.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -!- SIMILARITY: SOME, TO P.DENITRIFICANS COBW.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z74410; E249458; -
DR HYPOTHETICAL PROTEIN.
KW SEQUENCE 398 AA; 43699 MW; DAD8A665 CRC32;

Query Match 15.9%; Score 75; DB 1; Length 398;
Best Local Similarity 38.1%; Pred. No. 4.32e+00;
Matches 16; Conservative 7; Mismatches 16; Indels 3; Gaps 3;

Db 261 GVVTRTGRRLANR-PDQVMWLESAGGGLRVASAGKWLAA 301
QY 10 GVVKANGLLILQTRKPTGNHLKTSKG-M-VTWYKKWLLMT 49

Search completed: Wed Dec 22 10:26:35 1999
Job time : 8 secs.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:26:52 1999; MasPar time 10.53 Seconds
Tabular output not generated. 305.877 Million cell updates/sec
Title: >US-09-209-961-20
Description: (1-59) from US09209961.pep
Perfect Score: 471
Sequence: 1 MYTSLWHLGVVKGANGLLIL.....MVKKWLLMTFMAGCRGMII 59
Scoring table: PAM 150
Gap 11
Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries
Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus
Statistics: Mean 36.449; Variance 61.032; scale 0.597
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.
1 471 100.0 59 14 O92288 P6.6 3.89e-86
2 95 20.2 593 2 Q49695 B1496.F1.14. 6.74e-03
3 94 20.0 591 2 O53150 HYPOTHETICAL 62.7 KD P 9.89e-03
4 89 18.9 147 5 Q17666 COSMID C05E11. 6.49e-02
5 88 18.7 607 2 P96389 YDAA PROTEIN. 9.40e-02
6 83 17.6 387 6 Q46657 PROTAGLANDIN EP3 RECE 5.73e-01
7 82 17.4 129 5 Q95034 SUPEROXIDE DISMUTASE (8.16e-01
8 82 17.4 198 5 Q27740 SUPEROXIDE DISMUTASE (8.16e-01
9 80 17.0 101 14 Q68146 VPR PROTEIN. 1.64e+00
10 80 17.0 373 1 Q26344 SENSORY TRANSDUCTION H 1.64e+00
11 80 17.0 691 2 O54098 PUTATIVE MEMBRANE PROT 1.64e+00
12 80 17.0 885 2 P74259 HYPOTHETICAL 101.5 KD 1.64e+00
13 79 16.8 60 8 Q36334 UNIQUE ORF. 2.31e+00
14 79 16.8 578 5 O17444 VESICULAR ACETYLCHOLIN 2.31e+00
15 78 16.6 286 2 Q46525 FIMP. 3.26e+00
16 78 16.6 393 4 O00326 PROTAGLANDIN EP3 RECE 3.26e+00
17 78 16.6 394 4 O15192 EP3-III PROTEIN. 3.26e+00
18 78 16.6 402 4 O00325 PROTAGLANDIN EP3 RECE 3.26e+00
19 78 16.6 403 4 O15193 EP3-IV PROTEIN. 3.26e+00
20 78 16.6 419 4 O15549 EP3-I PROTEIN. 3.26e+00

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed Dec 22 10:26:52 1999; MasPar time 10.53 Seconds
Tabular output not generated. 305.877 Million cell updates/sec

Title: >US-09-209-961-20
Description: (1-59) from US09209961.pep
Perfect Score: 471
Sequence: 1 MYTSLWHLGVVKGANGLLIL.....MVKKWLLMTFMAGCRGMII 59

Scoring table: PAM 150
Gap 11

Searched: 179066 seqs, 54579741 residues
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrembl9
1:sp_archaea 2:sp_bacteria 3:sp_fungi 4:sp_human
5:sp_invertebrate 6:sp_mammal 7:sp_mhc 8:sp_organelle
9:sp_phase 10:sp_plant 11:sp_rodent 12:sp_unclassified
13:sp_vertebrate 14:sp_virus

Statistics: Mean 36.449; Variance 61.032; scale 0.597
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES
Result No. Score Query Match Length DB ID Description Pred. No.
1 471 100.0 59 14 O92288 P6.6 3.89e-86
2 95 20.2 593 2 Q49695 B1496.F1.14. 6.74e-03
3 94 20.0 591 2 O53150 HYPOTHETICAL 62.7 KD P 9.89e-03
4 89 18.9 147 5 Q17666 COSMID C05E11. 6.49e-02
5 88 18.7 607 2 P96389 YDAA PROTEIN. 9.40e-02
6 83 17.6 387 6 Q46657 PROTAGLANDIN EP3 RECE 5.73e-01
7 82 17.4 129 5 Q95034 SUPEROXIDE DISMUTASE (8.16e-01
8 82 17.4 198 5 Q27740 SUPEROXIDE DISMUTASE (8.16e-01
9 80 17.0 101 14 Q68146 VPR PROTEIN. 1.64e+00
10 80 17.0 373 1 Q26344 SENSORY TRANSDUCTION H 1.64e+00
11 80 17.0 691 2 O54098 PUTATIVE MEMBRANE PROT 1.64e+00
12 80 17.0 885 2 P74259 HYPOTHETICAL 101.5 KD 1.64e+00
13 79 16.8 60 8 Q36334 UNIQUE ORF. 2.31e+00
14 79 16.8 578 5 O17444 VESICULAR ACETYLCHOLIN 2.31e+00
15 78 16.6 286 2 Q46525 FIMP. 3.26e+00
16 78 16.6 393 4 O00326 PROTAGLANDIN EP3 RECE 3.26e+00
17 78 16.6 394 4 O15192 EP3-III PROTEIN. 3.26e+00
18 78 16.6 402 4 O00325 PROTAGLANDIN EP3 RECE 3.26e+00
19 78 16.6 403 4 O15193 EP3-IV PROTEIN. 3.26e+00
20 78 16.6 419 4 O15549 EP3-I PROTEIN. 3.26e+00

21 78 16.6 422 4 O15190 EP3E PROTEIN. 3.26e+00
22 73 16.6 422 4 O15189 EP3-VI PROTEIN. 3.26e+00
23 78 16.6 431 4 O15188 EP3-V PROTEIN. 3.26e+00
24 78 16.6 445 8 O79555 NADH DEHYDROGENASE SUB 3.26e+00
25 78 16.6 454 4 O15191 EP3F PROTEIN. 3.26e+00
26 78 16.6 535 3 O01165 TRANSPOSASE. 3.26e+00
27 78 16.6 692 2 Q45772 OUTER MEMBRANE PROTEIN 3.26e+00
28 78 16.6 792 2 O83999 CATION-TRANSPORTING AT 3.26e+00
29 77 16.3 142 2 Q44529 POTENTIAL MOLYBDENUM-P 4.57e+00
30 77 16.3 208 8 O21609 NADH DEHYDROGENASE SUB 4.57e+00
31 77 16.3 311 2 O65393 NAC GENE. 4.57e+00
32 77 16.3 359 5 O15839 L4171.5. 4.57e+00
33 77 16.3 434 1 O57980 434AA LONG HYPOTHETICA 4.57e+00
34 77 16.3 444 3 Q00921 MOX METHANOL OXIDASE. 4.57e+00
35 77 16.3 473 2 O24723 HYPOTHETICAL 49.1 KD P 4.57e+00
36 77 16.3 1073 5 O21885 COSMID R09H3. 4.57e+00
37 76 16.1 173 8 O79712 NADH DEHYDROGENASE SUB 6.39e+00
38 76 16.1 224 8 Q95896 NADH DEHYDROGENASE SUB 6.39e+00
39 76 16.1 328 3 O59938 FAMILY F XLANASE. 6.39e+00
40 76 16.1 448 10 O23100 BAC TM018A10. 6.39e+00
41 76 16.1 595 13 Q91751 FURIN (FRAGMENT). 6.39e+00
42 76 16.1 609 2 O84866 PROTEASE CONTAINING IR 6.39e+00
43 76 16.1 690 2 O33377 TRANSFERRIN-BINDING PR 6.39e+00
44 76 16.1 743 10 O82302 T32F12.27 PROTEIN. 6.39e+00
45 76 16.1 1007 6 O46432 LYOSOMAL ALPHA-MANNOS 6.39e+00

ALIGNMENTS

RESULT 1
ID O92288 PRELIMINARY; PRT; 59 AA.
AC O92288;
DT 01-NOV-1998 (TREMBLREL. 08, CREATED)
DT 01-NOV-1998 (TREMBLREL. 08, LAST SEQUENCE UPDATE)
-DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE P6.6.
CS PORCINE CIRCOVIRUS.
OC VIRUSES; SSDNA VIRUSES; CIRCOVIRIDAE; CIRCOVIRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-41:
RA WANG L., WILLSON P., CHOW B., GIBBONS E., BABIUK L.;
RT "Emergence of a new porcine circovirus."
RL SUBMITTED (AUG-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR ENBL; AF085695; G3668364;
SQ SEQUENCE 59 AA; 6393 MW; A40B778A CRC32;

Query Match 100.0%; Score 471; DB 14; Length 59;
Best Local Similarity 100.0%; Pred. No. 3.89e-86;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MYTSLWHLGVVKGANGLLILQTRPHTGNHLKTSGGVMTYVKWLLMTFMAGCRGMII 59
QY 1 MYTSLWHLGVVKGANGLLILQTRPHTGNHLKTSGGVMTYVKWLLMTFMAGCRGMII 59

RESULT 2
ID Q49695 PRELIMINARY; PRT; 593 AA.
AC Q49695;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
-DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE B1496.F1.14.
GN MLC1536.33.
OS MYCOBACTERIUM LEPRAE.
OC BACTERIA; FIRMICUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; CORYNEBACTERIACEAE; MYCOBACTERIACEAE; MYCOBACTERIUM.
RN [1]
RP SEQUENCE FROM N.A.
RA ROBISON K.;
SUBMITTED (NOV-1993) TO EMBL/GENBANK/DBJ DATA BANKS.
KJ [2]
SQ SEQUENCE FROM N.A.

Matches	-16;	Conservative	6;	Mismatches	18;	Indels	1;	Gaps	17;
---------	------	--------------	----	------------	-----	--------	----	------	-----

Db	70	LNLPSRI-QTVSLTMTTTGAVMMALAWMLGLRFTLGRRM	109
		:	
Qy	17	LLILOTRKPHGTGNHLSGGMVTMVKKLLLTTFMAGCRGM	57

RESULT		4		
ID	Q17666	PRELIMINARY;	PRT;	147 AA.
AC	Q17666;			
DT	01-NOV-1996	(TREMBLREL. 01, CREATED)		
DT	01-NOV-1996	(TREMBLREL. 01, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)		
DE	COSMID C05E11.			
GN	C05E11.2			

[1]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RC MEDLINE; 94150718;
RX
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA CONNELL M., COPESTY T., COOPER J., COULSON A.
RA CRAXTON M., BEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
OC RAABDITIDAE: RHABDITOIDEA: RHABDITIDAE: PELODERINAE: CAENORHABDITIS

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J.; PERCY C., RIFKEN L. ROOPRA A., SAUNDERS D., SHOWKNEEN

RA THERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA SHALDON N., SMITH A., SONNHÄTER E., STADEN R., SÜLSION J.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RL	NATURE 368:32-38(1994).
RN	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN-BRISTOL N2;
RC	STRAIN-BRISTOL N2;
RA	GEISEL C.;
RA	SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
RL	[3]
RN	SEQUENCE FROM N.A.
RP	STRAIN-BRISTOL N2;
RC	STRAIN-BRISTOL N2;
RA	WATERSTON R.;
RA	SUBMITTED (APR-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
DR	EMBL; U5338; G1255846; -
SQ	SEQUENCE 147 AA; 17070 MW; 037CD229 CRC32;
	Query Match 18.9%; Score 89; DB 5; Length 147;
	Best Local Similarity 27.9%; Pred. No. 6.49e-02;
	Matches 12; Conservative 14; Mismatches 16; Indels 1;

Dd	10	LLFALVSRPEASPETEFVVGVIVLKLIRKWLKRTFLARRGVV	52
		: : .: : : : : : :	
QY	17	LLILOTRKHITGNHLETSGGMTVMVK-KWLLLLTFMAGCRGMI	58
RESULT	5		
ID	96589	PRELIMINARY;	PRT; 607 AA.
AC	96589;		
DT	01-MAY-1997 (TREMREL. 03, CREATED)		
DT	01-JAN-1998 (TREMREL. 05, LAST SEQUENCE UPDATE)		
DT	01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)		
DE	YDQC PROTEIN.		
DE	YDAO.		
GN	BACILLUS SUBTILIS.		
OS	BACTERIA; FIRMICUTES; BACILLUS/CLOSTRIDIUM GROUP; BACILLACEAE;		
OC	BACILLUS.		
OC	[1]		
RN	SEQUENCE FROM N.A.		
RP			

RC SIRAIN-100;
RX MEDLINE: 98000887.
RA BELOIN C., AYORA S., EXLEY R., HIRSCHBEIN L., OGASAWARA N.,

DR EMBL; AJ001201; EL237902; -
SQ SEQUENCE 387 AA; 42758 MW; AE3DBEB6 CRC32;

Query Match 17.6%; Score 83; DB 6; Length 387;
Best Local Similarity 34.4%; Pred. No. 5.73e-01;
Matches 11; Conservative 9; Mismatches 11; Indels 1; Gaps 1;

Db 156 LAJRAPRSHYSSHKMTSATRAVLGLVWLAFLAF 187
QY 20 LQTRKPH-TGNHLKTSGGVTWVKWLLMTF 50

RESULT 7
ID Q95034 PRELIMINARY; PRT; 129 AA.
AC Q95034;
DT 01-FEB-1997 (TREMELREL_02, CREATED)
DI 01-FEB-1997 (TREMELREL_02, LAST SEQUENCE UPDATE)
DT 01-NOV-1996 (TREMELREL_08, LAST ANNOTATION UPDATE)
DE SUPEROXIDE DISMUTASE (EC 1.15.1.1) (FRAGMENT).
OS PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCR3;
RX MEDLINE; 97077434.
RA BECUPE P., GRATEPANCHE S., FOURMAUX M., VAN BEUMEN J., SAMYN B.,
RA MERCEREAU-PULALON O., TOUZEL J., SLOMIANNY C., CAMUS D., DIVE D.;
RT "Characterization of iron-dependent endogenous superoxide dismutase
of Plasmodium falciparum." ;
RL MOL. BIOCHEM. PARASITOL. 76:125-134(1996).
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) +
H2O(2).
CC -1- COFACTORS: COPPER AND ZINC OR IRON OR MANGANESE.
DR EMBL; Z25387; E85185; 1.
PFAM; PF00081; sofat; 1.
KW OXIDOREDUCTASE.
FT NON_TER. 1
FT NON_TER. 129
SQ SEQUENCE 129 AA; 14395 MW; 11B4898A CRC32;

Query Match 17.4%; Score 82; DB 5; Length 129;
Best Local Similarity 34.1%; Pred. No. 8.16e-01;
Matches 15; Conservative 11; Mismatches 15; Indels 3; Gaps 2;

Db 88 FGSGWGWLNNNNKLVILOTH--DAGNIKONTGPILTCDIW 129
QY 2 YTSLWGLGVKANGULLIRPHTGNHLKTSGGMVTV-KKW 44

RESULT 8
ID Q27740 PRELIMINARY; PRT; 198 AA.
AC Q27740;
DT 01-NOV-1996 (TREMELREL_01, CREATED)
DI 01-NOV-1996 (TREMELREL_01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMELREL_08, LAST ANNOTATION UPDATE)
DE SUPEROXIDE DISMUTASE (EC 1.15.1.1).
OS SOD.
GN GN
OC PLASMODIUM FALCIPARUM.
OC EUKARYOTA; ALVEOLATA; APICOMPLEXA; HAEMOSPORIDA; PLASMODIUM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HB3;
RX MEDLINE; 97077434.
RA BECUPE P., GRATEPANCHE S., FOURMAUX M., VAN BEUMEN J., SAMYN B.,
RA MERCEREAU-PULALON O., TOUZEL J., SLOMIANNY C., CAMUS D., DIVE D.;
RT "Characterization of iron-dependent endogenous superoxide dismutase
of Plasmodium falciparum." ;
RL MOL. BIOCHEM. PARASITOL. 76:125-134(1996).
CC -1- FUNCTION: DESTROYS RADICALS WHICH ARE NORMALLY PRODUCED WITHIN
THE CELLS AND ARE TOXIC TO BIOLOGICAL SYSTEMS.
CC -1- CATALYTIC ACTIVITY: 2 PEROXIDE RADICAL + 2 H(+) = O(2) +
H2O(2).
CC -1- SIMILARITY: BELONGS TO THE IRON/MANGANESE SUPEROXIDE DISMUTASE

```

SQ  SEQUENCE 373 AA; 42263 MW; A635962D CRC32;

Query Match 17.0%; Score 80; DB 1; Length 373;
Best Local Similarity 37.5%; Pred. No. 1.64e+00;
Matches 18; Conservative 6; Mismatches 20; Indels 4; Gaps 2;

Db 76 WGVVLVAVFVALLIQLPMTFVAG--LNISDELFRVLV--LLLLTFMVA 119
||| : : ||||| | : | : | ||| ||| :
QY 6 WGHZGVVYKANGLLIQLQPKPHTGNHLKTSGGVMTVMVKKWLLMTFMAG 53
||| : : ||||| | : | : | ||| ||| :

RESULT 11
ID O54098 PRELIMINARY; PRT; 691 AA.
AC O54098;
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE PUTATIVE MEMBRANE PROTEIN.
GN SC10A5.05
OS STREPTOMYCES COELICOLOR.
OC BACTERIA; FRMICTUTES; ACTINOBACTERIA; ACTINOBACTERIDAE;
OC ACTINOMYCETALES; STREPTOMYCINAE; STREPTOMYCETACEAE; STREPTOMYCES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA MURPHY L., HARRIS D.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA PARKHILL J., BARRELL B.G., RAJANDREAM M.A.;
RL SUBMITTED (JAN-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA REDENBACH M., KIESER H.M., DENAPATE D., EICHNER A., CULLUM J.,
RX MEDLINE; 97000351.
RA KINASHI H., HOPWOOD D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb. Streptomyces coelicolor A3(2) chromosome.";
RL MOL. MICROBIOL. 21:77-96(1996).
DR EMBL; AL021529; E1247672; -
SQ SEQUENCE 691 AA; 75164 MW; C92968B4 CRC32;

Query Match 17.0%; Score 80; DB 2; Length 691;
Best Local Similarity 29.5%; Pred. No. 1.64e+00;
Matches 13; Conservative 13; Mismatches 16; Indels 2; Gaps 2;

Db 195 GLVLDDTMRAPTADYEIKPEHGLAGFALIFLLLRFSFGCAAL 238
||| : : ||| : | : ||| : ||| :
QY 16 GLLIQLQPKPHTGNH-LKTS-GGVMTVMVKKWLLMTFMAGCRGM 57
||| : : ||||| | : | : | ||| ||| :

RESULT 12
ID P74259 PRELIMINARY; PRT; 885 AA.
AC P74259;
DT 01-FEB-1997 (TREMBLREL. 02, CREATED)
DT 01-FEB-1997 (TREMBLREL. 02, LAST SEQUENCE UPDATE)
DT 01-JAN-1999 (TREMBLREL. 09, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 101.5 KD PROTEIN.
OS SYNECHOCYSTIS SP. (STRAIN PCC 6803).
OC BACTERIA; CYANOBACTERIA; CHROCOCCALES; SYNECHOCYSTIS.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RA TABATA S.;
RL SUBMITTED (JUN-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-PCC6803;
RX MEDLINE; 97061201.
RA MIYAJIMA N., HIROSAWA M., KOTANI H., TANAKA A., ASAMIZU E., NAKAMURA Y.,
RA KANEKO T., SATO S., SUGIURA M., SASAMOTO S., KIMURA T.
```

RA HOSOUCHI T., MATSUNO A., MURAKI A., NAKAZAKI N., NARUO K., OKUMURA S.,
RA SHIMPO S., TAKEUCHI C., WADA T., WATANABE A., YAMADA M., YASUDA M.,
RA TABATA S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. PCC6803. II. Sequence determination of the entire
RT genome and assignment of potential protein-coding regions.";
RL DNA RES. 3:109-136(1996).
DR EMBL; D90913; D1019086; -;
DR PFAM; PF00128; alpha-amylase; 1.
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 885 AA; 101491 MW; C9C80BAE CRC32;

Query Match 17.0%; Score 80; DB 2; Length 885;
Best Local Similarity 39.5%; Pred. No. 1.64e+00;
Matches 15; Conservative 7; Mismatches 13; Indels 3; Gaps 3;
Db 409 MYGODLNHOLPIVRAI-LLEMORRKINTGADGIRVDGG 445
QY 1 MYT-SLGHGLGVKANGLLILQTRKPHGNH-LKTSGG 36

RESULT 13
ID Q36334 PRELIMINARY; PRT; 60 AA.
AC Q36334;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE UNIQUE ORF.
GN PUTATIVE ORF60.
OS CHONDRUS CRISPUS (CARRAGHEEN).
OG MITOCHONDRION.
OC EUKARYOTA; RHODOPHYTA; FLORIDOPHYCEAE; GIGARTINALES; GIGARTINACEAE;
OC CHONDRUS.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-APICES;
RX MEDLINE; 95341681.
RA LEBLANC C., BOYEN C., RICHARD O., BONNARD G., GRIENENBERGER J.M.,
RA KLOAREG B.;
RT "Complete sequence of the mitochondrial DNA of the rhodophyte
RT Chondrus crispus (Gigartinales). Gene content and genome
RT organization.";
RL J. MOL. BIOL. 250:484-495(1995).
DR EMBL; Z47547; G1019071; -;
KW MITOCHONDRION.
SQ SEQUENCE 60 AA; 6633 MW; 49A006C5 CRC32;

Query Match 16.8%; Score 79; DB 8; Length 60;
Best Local Similarity 42.9%; Pred. No. 2.31e+00;
Matches 6; Conservative 7; Mismatches 1; Indels 0; Gaps 0;
Db 1 MYRHWLLMLAFKGS 14
QY 40 MYRHWLLMLTFMAG 53

RESULT 14
ID O17444 PRELIMINARY; PRT; 578 AA.
AC O17444;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-JAN-1998 (TREMBLREL. 05, LAST ANNOTATION UPDATE)
DE VESICULAR ACETYLCHOLINE TRANSPORTER.
GN VACHT.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A.
RA KITAMOTO T., WANG W., SALVATERRA P.M.;
RL J. BIOL. CHEM. 0:0-0(1997).
DR EMBL; AF030197; G2625056; -;

SQ SEQUENCE 578 AA; 64345 MW; 4AE4211D CRC32;
Query Match 16.8%; Score 79; DB 5; Length 578;
Best Local Similarity 31.8%; Pred. No. 2.31e+00;
Matches 14; Conservative 9; Mismatches 18; Indels 3; Gaps 3;
Db 22H DGLMLLLVMPKVEA-MKOSKDVQVQVPIWRLLMDPYAVACAG 270
QY 15 NGLLIQTRKPHGTGNHKTSGMVT-MVRKKWLLLM-TFMAGCRG 56

RESULT 15
ID Q46525 PRELIMINARY; PRT; 286 AA.
AC Q46525;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE FIMP.
GN FIMP.
OS BACTERIOIDES NODOSUS (DICHELOBACTER NODOSUS).
OC BACTERIA; PROTEOBACTERIA; GAMMA SUBDIVISION; CARDIOBACTERIACEAE;
OC DICHELOBACTER.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A198;
RX MEDLINE; 95369727.
RA JOHNSTON J.L., BILLINGTON S.J., HARING V., ROOD J.I.;
RT "Identification of fibrial assembly genes from Dichelobacter
RT nodosus: evidence that fimp encodes the type-IV prepeilin peptidase.";
RL GENE 161:21-26(1995).
DR EMBL; U17138; G790211; -;
SQ SEQUENCE 286 AA; 31863 MW; EA572E0E CRC32;

Query Match 16.6%; Score 78; DB 2; Length 286;
Best Local Similarity 28.8%; Pred. No. 3.26e+00;
Matches 12; Conservative 14; Mismatches 14; Indels 2; Gaps 2;
Db 67 LLVPASRCPHCGRIRAIEN-IPVIS-WLFKKGKSCGGAAI 106
QY 17 LLILQTRKPHGTGNHKTSGMVT-MVRKKWLLLM-TFMAGCRGMI 58

Search completed: Wed Dec 22 10:27:07 1999
Job time : 15 secs.

(TM)

Result No.	Score	Query Match	Length	ID	Description	Pred. No.
1	81	19.5	317 28	W33721	Human melancortin-1	3.82e+01
2	81	19.5	317 36	W79684	Melanocortin-1 recept	3.82e+01
3	81	19.5	317 22	W19706	Melanocortin-1 recept	3.82e+01
4	81	19.5	317 9	R49725	Sequence of a polypep	3.82e+01
5	81	19.5	317 8	R43571	Human MSH-R.	3.82e+01
6	80	19.2	317 31	W78327	Human melanocyte stim	4.53e+01
7	79	19.0	246 36	W77312	Polyptide sequence	5.37e+01
8	79	19.0	404 31	W59035	Human G-protein coupl	5.37e+01
9	79	19.0	404 28	W31344	Human G-protein coupl	5.37e+01
10	79	19.0	404 28	W34984	Human G-protein coupl	5.37e+01
11	79	19.0	1099 36	W29683	Human Na ⁺ -K ⁺ -2Cl cotran	5.37e+01
12	78	18.8	197 34	W77749	Staphylococcus aureus	6.36e+01
13	78	18.8	400 13	R71932	I-branching enzyme.	6.36e+01
14	78	18.8	400 16	R92474	Beta-1,6-N-acetylgluc	6.36e+01
15	78	18.8	400 29	W47184	Human beta-1,6-N-acet	6.36e+01
16	78	18.8	400 31	W56622	Full length amino aci	6.36e+01

AC W79684;
 DT 17-DEC-1998 (first entry)
 DE Melanocortin-1 receptor.
 KW Human melanocortin-1 receptor; MC1; homology; transmembrane domain;
 KW G-protein; MC3; therapeutic.
 OS Homo sapiens.
 PN US5817787-A.
 PD 06-OCT-1998.
 PF 23-APR-1997; 842045.
 PR 17-FEB-1994; US-200711.
 PR 27-JUN-1996; US-672109.
 PR 23-APR-1997; US-842045.
 PA (UNMI) UNIV MICHIGAN.
 PI Gantz I, Yamada T;
 DR N-PSDB; V62349.
 DR N-PSDB; V62349.
 PT DNA encoding melanocortin-5 receptor - useful in hybridisation
 PT assays for melanocortin-5 receptor nucleic acids
 PS Disclosure; Column 29-32; 58pp; English.
 CC The present sequence represents the human melanocortin-1 (MC1) receptor,
 CC the gene of which has been localised to chromosome 16q24.3. To obtain
 CC the DNA sequence, RNA was first isolated from HeLa cells and U937 cells
 CC and then subjected to RT-PCR, using primers based on the homologous
 CC regions of the third and sixth transmembrane domains of G-protein linked
 CC receptors. This sequence as well as MC3 (V62351), were then used to
 CC construct oligonucleotides that could be used to search genomic DNA for
 CC other members of the receptor family. These genes and their products may
 CC be used to provide therapeutic vehicles for the treatment of processes
 CC involving the function of melanocortin receptors.
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 36; Length 317;
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 w-vasvftlfiaydhvavllclvfvf 195
 QY 11 WGVFKIKFSELYIHGYTDIVLVVFTVF 38

RESULT 3
 ID W19706 standard; Protein; 317 AA.
 AC W19706;
 DT 19-AUG-1997 (first entry)
 DE Melanocortin-1 receptor.
 KW Melanocortin-3; MC3R; MC1R; MC2R; MC4R; MC5R; human; mouse; melanocyte;
 KW pro-opiomelanocortin; adrenal cortical function; behaviour; learning;
 KW memory; cardiovascular system; analgesia; thermoregulation; prolactin;
 KW neurohumoral agent; biogenic amine.
 OS Homo sapiens.
 PN US5622860-A.
 PD 22-APR-1997.
 PF 17-FEB-1994; 200711.
 PR 17-FEB-1994; US-200711.
 PA (UNMI) UNIV MICHIGAN.
 PI Gantz I, Yamada T;
 DR N-PSDB; T68792.
 DR Nucleic acid molecules encoding melanocortin receptors - useful to
 PT transfect mammalian cells lacking endogenous receptors to induce
 PT their expression
 PS Example 1; Column 29-32; 58pp; English.
 CC This sequence represents the human and mouse melanocortin (MC) receptors.
 CC W19703-W19707 represent the human and the gene encoding it is located at
 CC chromosome locus 16q24.3. MCs are products of pro-opiomelanocortin
 CC post-translational processing, and are known to have a broad array of
 CC physiological actions. MCs are known to have effects on adrenal cortical
 CC functions and on melanocytes, as well as affecting behaviour, learning,
 CC memory, control of the cardiovascular system, analgesia, thermoregulation
 CC and the release of other neurohumoral agents (such as prolactin and
 CC biogenic amines). The nucleic acids can be used to transfect mammalian
 CC cells lacking endogenous MC receptors to induce their expression. These
 CC sequences can also be used to screen and identify drugs which

CC specifically react with MCRs on the surface of a cell. The drugs can then
 CC be used for treating diseases which have MCRs implicated as one of their
 CC causes. Vectors containing these sequences can also be used to treat the
 CC diseases.
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 22; Length 317;
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 w-vasvftlfiaydhvavllclvfvf 195
 QY 11 WGVFKIKFSELYIHGYTDIVLVVFTVF 38

RESULT 4
 ID R49725 standard; Protein; 317 AA.
 AC R49725;
 DT 20-AUG-1994 (first entry)
 DE Sequence of a polypeptide (PP) with the properties of
 DE melanotropic hormone receptor
 DE Melanocyte stimulating hormone receptor
 KW Melanocyte stimulating hormone receptor; melanotropic; cancer;
 KW diagnosis; therapy.
 OS Homo sapiens.
 PN WO9404674-A.
 PD 03-MAR-1994.
 PF 20-AUG-1993; DK0273.
 PR 21-AUG-1992; DK-001046.
 PR 10-SEP-1992; DK-001118.
 PR 05-MAY-1993; DK-000528.
 PA (CHHA/) CHHAJLANI V.
 PA (WIKB/) WIKBERG J.
 PI Chhajlani V, Wikberg J;
 DR WPI; 94-083193/10.
 DR N-PSDB; Q44340.
 PT Novel DNA fragments encoding novel peptides - having properties
 PT of melanotropic hormone receptors esp. encoding melanocyte
 PT stimulating hormone receptors, are used to diagnose e.g. skin
 PT cancers or vitiligo
 PS Claim 10; Page 95-97; 134pp; English.
 CC The DNA fragment with the sequence in Q44340 is derived from a human
 CC cDNA library. The encoded polypeptide (R49725) constitutes the
 CC entire polypeptide of a melanocyte stimulating hormone (MSH)
 CC receptor. The mRNA of this clone was found only in the melanoma
 CC cells and not in other tissues examined including brain, thymus,
 CC parathyroid gland, parotid gland, salivary gland, adrenal gland,
 CC testis, liver, lung, heart, spleen, skeletal muscle, intestine and
 CC colon. Transmembrane sites in the polypeptide corresp. to bps 286-
 CC 351, 394-465, 517-588, 640-711, 733-804, 898-972, 997-1088.
 CC Glycosylation sites are found at AA residues 15 and 29; possible
 CC phosphorylation sites are found at AA residues 42-45, 151-154,
 CC 306-308; and a possible palmitoylation site is found at AA
 CC residue 316.
 SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 9; Length 317;
 Best Local Similarity 39.3%; Pred. No. 3.82e+01;
 Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 w-vasvftlfiaydhvavllclvfvf 195
 QY 11 WGVFKIKFSELYIHGYTDIVLVVFTVF 38

RESULT 5
 ID R43571 standard; Protein; 317 AA.
 AC R43571;
 DT 22-MAY-1994 (first entry)
 DE Human MSH-R.
 KW Melanocyte stimulating hormone receptor; mouse; genetic disease;
 KW isolation; screening; antagonists.
 OS Homo sapiens.
 PN WO9321316-A.

```

PD 28-OCT-1993. U03247.
PF 07-APR-1993; US-866979.
PR 10-APR-1992; US-866979.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Cone RD, Mountjoy KG;
DR WPI: 93-351726/44.
DR N-PSDB; Q51226.
PT Melanocyte stimulating hormone receptors and corresp. DNA - from
PT humans and mice, used for diagnosis of genetic disease and for
PT screening agonists
PS Claim 8; Fig 1B; 46pp; English.
CC The sequence is that of the human melanocyte-stimulating hormone
CC receptor isolated from a human melanoma cDNA library using a pair
CC of degenerate PCR primers corresp. to the putative third and sixth
CC transmembrane regions of G-protein coupled receptors. The sequence
CC can be used as a probe for the detection and diagnosis of genetic
CC diseases, or for the isolation of novel mammalian receptors and genes.
CC It may also be used for screening and quantitating cpds. as inhibitors
CC of agonist binding to the receptor.
CC See also R43571.
SQ Sequence 317 AA;

Query Match 19.5%; Score 81; DB 8; Length 317;
Best Local Similarity 39.3%; Pred. No. 3.82e+01;
Matches 11; Conservative 5; Mismatches 11; Indels 1; Gaps 1;

Db 169 w-vasvfvstlfi-gyydhvavllclvlf 195
QY 11 WGVFKIFSELYIHGYTDIVLVVFTVF 38

RESULT 6
ID W37827 standard; Protein; 317 AA.
AC W37827;
DT 28-JUL-1998 (first entry)
DE Human melanocyte stimulating hormone-1 receptor.
KW Human melanocyte stimulating hormone-1 receptor gene; MCRL;
KW MCRL; agonist; antagonist; feeding; metabolic disorder;
KW eating disorder; anorexia; obesity; cachexia; cancer; inhibition.
OS Homo sapiens.
PN W09810068.A2.
PD 12-MAR-1998.
PF 04-SEP-1997; U15565.
PR 04-SEP-1996; US-706281.
PA (UYOR-) UNIV OREGON HEALTH SCI.
PI Boston BA, Chen W, Cone RD, Fan W, Kesterton RA,
PI Lu D;
DR WPI: 98-193618/17.
DR N-PSDB; V19136.
PT Identifying melanocortin receptor agonists and antagonists - using a
PT panel of recombinant mammalian cells expressing alpha-melanocyte
PT stimulating hormone, ACTH, MC-3, MC-4 and MC-5 receptors
PS Example 2B; Fig 2A-B; 121pp; English.
CC This is the amino acid sequence of the human melanocyte stimulating
CC hormone receptor (MCRL). The MCR agonists and antagonists can be used
CC for modifying feeding behaviour in an animal. The antagonists can be
CC used for stimulating feeding while the agonists can be used for
CC inhibiting feeding. They can be used for the treatment of eating
CC disorders such as anorexia and obesity, and other pathological weight
CC and eating-related disorders. They can also be used to treat failure
CC to thrive disorders and disease-related cachexia, such as occurs in
CC cancer patients, as well as other metabolic disorders.
SQ Sequence 317 AA;

Query Match 19.2%; Score 80; DB 31; Length 317;
Best Local Similarity 44.8%; Pred. No. 4.53e+01;
Matches 13; Conservative 5; Mismatches 8; Indels 3; Gaps 3;

Db 169 w-vasvfvstlfi-gyydhvavllclvlf 195
QY 11 WGVFKIFSELYIHGYTDIVLVVFTVF 38

```

```

RESULT 7
ID W77312 standard; protein; 246 AA.
AC W77312;
DT 16-DEC-1998 (first entry)
DE Polypeptide sequence of p28 Bap31/CDM.
KW Apoptosis; p28 Bap31/CDM; modulate; p28 Bap31; human; p20 product;
KW pro-FUICE/Bcl-2 protein; Bcl-2; Bcl-XL; degenerative disease; neoplasia.
OS Homo sapiens.
FH Key.
FT Peptide
FT 2..11 Location/Qualifiers
FT /note= "sequence homologous to human Bap31"
FT Domain
FT 7..27 /note= "predicted transmembrane domain TM1"
FT Domain
FT 43..64 /note= "predicted transmembrane domain TM2"
FT Domain
FT 102..123 /note= "predicted transmembrane domain TM3"
FT Peptide
FT 161..165 /note= "caspase recognition site"
FT Cleavage_site
FT 164..165 /note= "caspase recognition site"
FT Peptide
FT 235..239 /note= "caspase recognition site"
FT Cleavage_site
FT 238..239 /note= "ER retention signal sequence"
FT Peptide
FT 243..246
FT /note= "ER retention signal sequence"
PN W09839434-AA.
PD 11-SEP-1998.
PF 02-MAR-1998; IB0706.
PR 03-MAR-1997; CA-198988.
PA (UYMC-) UNIV MCGILL.
PI Branton PE, Ng WHF, Nguyen M, Shore GC;
PI WPI: 98-503359/43.
PT Polypeptide-modulating apoptosis, p28 Bap31 - useful to diagnose
PT diseases involving altered apoptosis, e.g. degenerative diseases or
PT neoplasia and to detect compounds modulating apoptosis
PS Claim 7; Fig 2C; 130pp; English.
CC This represents the amino acid sequence of the p28 Bap31/CDM polypeptide
CC that can modulate apoptosis. The polypeptide p28 Bap31, fragments and
CC encoding nucleic acids are useful to generate or identify compounds which
CC alter the biological activity of p28 Bap31 and thus modulate apoptosis
CC when administered to cells (e.g. mammalian and especially human or rodent
CC cells). Such compounds may affect e.g. p28 Bap31 cleavage to produce p20
CC product, formation of complex with pro-FUICE/Bcl-2 protein (especially
CC Bcl-2 or Bcl-XL), specific binding to an antibody or p28 Bap31 expression
CC in the cell. The compounds are especially useful to inhibit apoptosis in
CC mammals with degenerative diseases or to increase apoptosis in mammals
CC with neoplasia. p28 Bap31 can also be used diagnostically, by measuring
CC levels in mammalian samples whereby a reduction is indicative of the
CC presence or likelihood of a disease caused by decreased apoptosis
CC (neoplasia) and an increase indicates the presence of a disease caused
CC by increased apoptosis (especially a degenerative disease). Antibodies
CC specifically binding p28 Bap31/p20 can similarly be used to diagnose
CC diseases/likelihood of diseases involving altered apoptosis.
SQ Sequence 246 AA;

Query Match 19.0%; Score 79; DB 36; Length 246;
Best Local Similarity 38.7%; Pred. No. 5.37e+01;
Matches 12; Conservative 3; Mismatches 8; Indels 2; Gaps 2;

Db 31 rwqkifsrllvlysgntffvllvllv 61
QY 10 WGVFKIFSELYIHGYTDIVLVVFTVF 38

RESULT 8
ID W59335 standard; protein; 404 AA.
AC W59335;
DT 04-AUG-1998 (first entry)
DE Human G-protein coupled receptor protein.
KW G-protein coupled receptor; human; cerebellum; adrenomedullin; asthma;
KW screening; agonist; antagonist; diagnosis; disease; osteoporosis;
KW Parkinson's disease; acute heart failure; urinary retention.
OS Homo sapiens.

```

Query Match 19.0%; Score 79; DB 28; Length 404;
Best Local Similarity 34.4%; Pred. No. 5.37e+01;
Matches 11; Conservative 9; Mismatches 3; Indels 3; Gaps 3;

Thiazide sensitive cotransporter, ATP sensitive K channel and NaK2C1 cotransporter genes - useful for developing products for the diagnosis and treatment of ion transport disorders, e.g. Gitelman's Syndrome or Bartter's Syndrome
 Disclosure; Fig.7; 105pp; English.
 This is the amino acid sequence of human renal bumetanide-sensitive

CC Na-k-2Cl cotransporter NKCC2 of the thick ascending limb of Henle's
 CC loop. Mutations in the NKCC2 gene (see V40562) cause Bartter's
 CC syndrome, a disorder featuring salt wasting and hypokalaemic
 CC alkalosis associated with marked hypercalcaemia and frequently
 CC nephrocalcinosis. Different molecular variants inferred to alter
 CC the structure of NKCC2 protein were identified in patients with
 CC Bartter's syndrome by SSCP using 27 pairs of specific primers (see
 CC V45677-730). Identification of the molecular basis of Bartter's
 CC syndrome provides for the genetic diagnosis of this disorder. A
 CC method is claimed for determining the presence or absence of a
 CC mutated protein conferring altered ion transport by analysing
 CC NKCC2 samples. Also claimed are vectors, host cells, a method of
 CC producing NKCC2 protein, and an antibody that binds altered, but
 CC not wild-type, NKCC2 protein. The products and methods can be used
 CC for diagnosis of Bartter's syndrome, to identify carriers and to
 CC identify modulators of NKCC2 function of therapeutic value.
 SQ Sequence 1099 AA;

Query Match 19.0%; Score 79; DB 36; Length 1099;
 Best Local Similarity 35.7%; Pred. No. 5.37e+01;
 Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 2;

Db 966 frikfadihiqinirpnkeskvfeemierpyrlhesckdl 1007

QY 14 FKIKSELVINGTDTI-VLVVTFVFSERAEAYVWHIS-RGL 53

RESULT 12

ID W77749 standard; Protein; 197 AA.

AC W77749;

DT 30-OCT-1998 (first entry)

DE Staphylococcus aureus protein of unknown function.

KW Staphylococcus aureus protein; immune response induction; eye infection;
 KW antibody production; T-cell immune response; gastrointestinal infection;
 KW respiratory infection; inhibitor; bacterial infection; cardiac infection;
 KW central nervous system; kidney infection; urinary tract infection;
 KW antimicrobial compound identification; broad spectrum antibiotic;
 KW therapy.

OS Staphylococcus aureus.

FH Key Location/Qualifiers

FT Misc_difference 1..197
 FT /note- "residues designated X are unspecified, and
 represented as Xaa in the specification"

FT EP-841394-A2.

PN 13-MAY-1998.

PD 24-SEP-1997; 307485.

PR 24-SEP-1996; US-027032.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Black MT, Burnham MKR, Hodgson JE, Knowles DJC,

PI Lonetto MA, Nicholas RO, Pratt JM, Reichard RW, Rosenberg M,

PI Ward JM;

DR WPI: 98-252940/23.

DR N-PSDB: V53537.

PT New nucleic acid sequences from Staphylococcus aureus WCHU29 -

PT useful in vaccines and for treatment of bacterial infections of e.g.

PT respiratory tract and central nervous system

PS Claim 11; Page 369-370; 390pp; English.

CC This sequence represents a Staphylococcus aureus protein of unknown
 CC function, and is encoded by a DNA sequence of the invention.

CC The DNA sequences were isolated from Staphylococcus aureus WCHU29

CC (NCIMB 40771). Host cells containing the DNA sequences are used to

CC produce polypeptides or fragments. The proteins are used in the treatment

CC of disease, for inducing an immune response by administering them, to

CC produce antibody and/or T-cell immune response. Antagonists of the

CC proteins are used for the inhibition of bacterial polypeptides.

CC Conditions which may be treated include bacterial infections, especially

CC respiratory, cardiac, gastrointestinal, central nervous, eye, kidney,

CC urinary tract, skin, bones and joints. The proteins can also be used to

CC identify antimicrobial compounds which are broad spectrum antibiotics,

CC especially useful in the treatment of H. pylori infection.

CC Sequence 197 AA;

Query Match 18.8%; Score 78; DB 34; Length 197;
 Best Local Similarity 16.2%; Pred. No. 6.36e+01;
 Matches 6; Conservative 16; Mismatches 13; Indels 2; Gaps 2;

Db 89 kixavvyq-yddttisif-vydvvnDewilrlqpn1 123

QY 17 KFSEYVHGTYDVLVVFVFSERAEAYVWHISRGL 53

RESULT 13

ID R71932 standard; Protein; 400 AA.

AC R71932;

DT 23-SEP-1995 (first entry)

DE I-branching enzyme.

KW Beta-1,6-N-acetylglucosaminyltransferase; I-branching enzyme; IGNT;

KW inflammation; antiinflammatory; tumor; hypersensitivity; anemia;

KW transgenic animal: I antigen.

OS Homo sapiens.

PN W09507020-A.

PD 16-MAR-1995.

PF 09-SEP-1993; U08476.

PA (LJOL-); LA JOLLA CANCER RES FOUND.

PI Bierhuizen MFA, Fukuda M;

DR WPI: 95-123182/16.

DR N-PSDB: Q89201.

PT Nucleic acid encoding I-branching enzyme - used to develop prods.

PT for treating e.g. inflammatory responses, tumours or

PT hypersensitivity reactions

PS Disclosure; Fig.6; 56pp; English.

CC Poly-A RNA isolated from human PA-1 teratocarcinoma cells was

CC reverse-transcribed and inserted into expression vector pcDNA1.

CC Plasmid DNA was used to transfect CHO-py-leu cells. Transfected

CC cells were screened using human anti-I antigen antibodies and goat

CC anti-human IgM. Plasmid DNA was analyzed to obtain DNA (given in

CC Q89201) encoding I-branching enzyme (R71932).

CC Sequence 400 AA;

Query Match 18.8%; Score 78; DB 13; Length 400;

Best Local Similarity 25.0%; Pred. No. 6.36e+01;

Matches 14; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

Db 8 lfliisvsvliifvsvfnfggdp 31

QY 21 LYIHGYTDIVLVVVFVFSERAE 44

RESULT 14

ID R92474 standard; Protein; 400 AA.

AC R92474;

DT 22-APR-1996 (first entry)

DE Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme.

KW Beta-1,6-N-acetylglucosaminyltransferase, I-branching enzyme; IGNT;

KW glycosyltransferase; blood group: I antigen; polylactosaminoglycan.

OS Homo sapiens.

FH Key Location/Qualifiers

FT domain 7:25

FT /label- Membrane-anchoring_domain

FT modified_site 37

FT /label- N-glycosylation_site

FT modified_site 212

FT /label- N-glycosylation_site

FT modified_site 255

FT /label- N-glycosylation_site

FT modified_site 314

FT /label- N-glycosylation_site

FT modified_site 388

FT /label- N-glycosylation_site

PN US5484593-A.

PD 16-JAN-1996.

PF 09-SEP-1993; 118906.

PR 09-SEP-1993; US-118906.

PA (LJOL-) LA JOLLA CANCER RES FOUND.


```
Query Match.      18.8%; Score 78; DB 29; Length 400;  
Best Local Similarity 25.0%; Pred. No. 6,36e+01;  
. Matches       6; Conservative 14; Mismatches 4; Indels  
  
db          8 lfisvssvilfivsfvnfggdp 31  
|.|.:|.:.|||.:.:  
QY         21 LYINGTYDIWLVVFTVFERSAEA 44
```

Search completed: Wed Dec 22 10:28:45 1999
Job time : 18 secs.